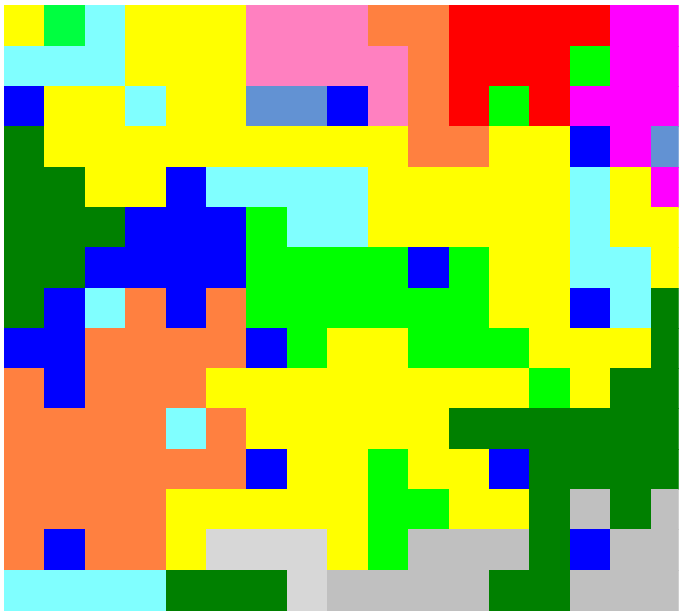
Data Uncertainty Engine (DUE)



User's Manual

James D. Brown

Institute for Biodiversity and Ecosystem Dynamics, Universiteit van Amsterdam, 1018 WV Amsterdam, The Netherlands, e-mail: brown@science.uva.nl

Gerard B.M. Heuvelink

Soil Science Centre, Wageningen University and Research Centre, P.O. Box 47, 6700 AA Wageningen, The Netherlands, e-mail: gerard.heuvelink@wur.nl

Date Uncertainty Engine (DUE), Version 3.1

Copyright © James D. Brown and Gerard B.M. Heuvelink

Data Uncertainty Engine (DUE) is free software; you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation; either version 2 of the License, or (optionally) any later version. DUE is distributed in the hope that it will be useful, but without any warranty; without even the implied warranty of merchantability or fitness for a particular purpose. See the GNU General Public License for more details. You should have received a copy of the GNU General Public License along with the program; if not, write to the Free Software Foundation, Inc., 675 Mass Ave, Cambridge, MA 02139, USA.

Contents

1. Introduction …… 5

2. Installation and start-up……. 7

2.1 Requirements……... 7

2.2 Unpacking and running DUE.……. 7

2.3 Troubleshooting the installation……..……..……...…….. 8

2.4 Altering memory settings……..…….……. 8

2.5 Source code and documentation…….. 9

3. Overview of functionality……..……. 10

3.1 Summary of functionality in DUE Version 3.1..……. 10

3.2 Planned functionality……..……. 11

4. Getting started……. 13

4.1 Performing an uncertainty analysis with DUE…… 13

4.2 Administrative functions……. 13

4.3 Importing and exporting data with files……..…… 15

4.4 Importing and exporting data with a DUE-enabled database…….…… 20

4.5 Creating projects…..……...……...……. 23

5. Examples and exercises…..…….. 24

5.1 Importing a time-series into DUE from file...……. 24

5.2 Importing a time-series into DUE from the prototype database……. 24

5.3 Defining an uncertainty model for a time-series …… 25

5.4 Defining a correlation model for an uncertain time-series…….. 30

5.5 Generating realisations of an uncertain time-series……. 33

5.6 Generating realisations of a spatial raster attribute with sample data……. 34

5.7 Generating realisations of spatial vector objects……. 37

5.8 Co-simulation of multiple cross-correlated time-series..……. 40

APPENDIX A1. The conceptual basis for DUE…… 43

A1.1 Introduction……..…..……. 43

A1.2 Objects and attributes……..……. 43

A1.3 A taxonomy of positional uncertainty……..…… 43

A1.4 A taxonomy of attribute uncertainty……...…….…… 44

APPENDIX A2. Models and algorithms used in DUE.…… 46

A2.1 Introduction……..…..……. 46

A2.2 Attribute uncertainty……..……. 46

A2.3 Positional uncertainty……...…..……. 50

A2.4 Simulation from probability models…… 50

APPENDIX A3. References…….…...…… 52

1. INTRODUCTION

Environmental models typically rely on input data, such as rainfall, flow boundary conditions, slope, terrain elevation and soil moisture, to make predictions about past, current or future states of the environment. In practice, the values of these inputs are rarely certain. Uncertainties may originate from imprecise measurements, sampling, interpolation, positional errors, and cartographic generalisation, among others. If the inputs of an environmental model are uncertain, the predictions will also be uncertain, because uncertainties propagate through a model. Other sources of uncertainty in model predictions include the structure, parameters and solution methods used. Together, these uncertainties can adversely affect policy or management decisions because the accuracy and precision of the model predictions is insufficient or poorly quantified.

The Data Uncertainty Engine (DUE) allows uncertainties in model inputs to be described and their impacts propagated through to model predictions. Sample data may be used alongside expert judgement to help construct an uncertainty model (UM) with DUE. Typically, sample data will improve the quality of a UM and may be used to: 1) help identify the parameters of the UM and; 2) to reduce the uncertainty of the simulated output by ensuring the realisations honour the sample data at some specified locations, as well as the UM itself.

A UM may be defined for one or more (possibly related) inputs using a probability distribution, a confidence interval or a set of possible outcomes (scenarios), depending on available knowledge and expertise. Uncertainty propagation is quantified by sampling from the uncertain inputs and implementing the model for each 'realisation' of the input values. In order to perform an uncertainty propagation analysis with DUE, realisations may be written to file and used in an external model. Alternatively, uncertainty models may be called programmatically from other software via a simple Application Programming Interface. Sensitivity analysis, parameter optimisation, data assimilation and assessments of structural uncertainty in models are not supported by DUE. While parameter optimisation is not allowed in DUE, parameter uncertainties can be treated in a similar way to other (e.g. measured) types of environmental variable and are, therefore, accommodated by DUE.

Using DUE, the spatial and temporal patterns of uncertainty (autocorrelation), as well as cross-correlations between related inputs, can be incorporated in an uncertainty analysis. Such correlations may greatly influence the outcomes of an uncertainty analysis because models typically respond differently to correlated variability than random errors. DUE also supports the quantification of positional uncertainties in geographic objects, represented as raster maps, time-series or vector outlines. Most importantly, DUE provides a conceptual framework for structuring an uncertainty

analysis, allowing user's without direct experience of statistical methods for uncertainty propagation to develop realistic UM for their data. As with more generic tools (e.g. R, SPLUS, Matlab), the quality of a UM will depend on the user's level of expertise and knowledge of the data, but unlike these tools, DUE provides a structured user interface and framework of assumptions (that must be justified) for constructing and estimating a UM given limited resources.

Data may be loaded into DUE from file or from a database, and are stored within DUE as objects, whose positions may be uncertain, and attributes, whose values may be uncertain. For attributes that vary continuously in space or time, such as terrain elevation, rainfall or river discharge, positional uncertainty leads to uncertainty in the attribute values, and can be incorporated as attribute uncertainty in DUE. Objects supported by DUE include spatial vectors, space-time vectors, spatial rasters, time-series of rasters, simple time-series and objects that are 'constant' in space and time. Attributes supported by DUE include continuous numerical variables (e.g. rainfall), discrete numerical variables (e.g. bird counts) and categorical variables (e.g. land-cover).

2. INSTALLATION AND START-UP

2.1 Requirements

In order to run DUE on a PC Workstation you will need:

1. The Java™ Runtime Environment (JRE) version 5.0 (1.5.0) or higher. The JRE is free software and may be downloaded from the Sun website:

   http://java.sun.com/j2se/index.jsp

2. The DUE executable, DUE.jar, and associated resources in DUE_3.1.zip;

3. Microsoft Windows 98/2000/NT/XP Operating System (OS). The software has not been tested on other OS but will be available for Linux, UNIX, Macintosh or other platforms shortly. On a Windows platform, you will need:

   - A minimum of 32MB of RAM and ~50MB of hard-disk space free.
   - For many practical applications of DUE, including simulation from large datasets (more than ~100,000 values), more RAM may be required. A minimum of 512MB is recommended.

4. External tools to visualise realisations of spatial or temporal datasets. Currently, many other proprietary and free software tools are available for data visualisation, such as Landserf (also written in Java and freely available from www.landserf.org).

2.2 Unpacking and running DUE

Once you have obtained the DUE software, unpack the zipped archive to any directory on your PC (e.g. C:/Program Files/DUE_3.1/) using WinZip™ or similar software. Do not move the DUE.jar executable from the existing directory structure: create a shortcut elsewhere if required.

Once you have unpacked the software, you may run DUE by double-clicking on "DUE.jar" or by navigating to the root directory and typing "./DUE.jar" in a command prompt. For access outside the installation directory, add a reference to "DUE.jar" in the system path (on Windows machines).

2.3 Troubleshooting the installation

List of typical problems and actions:

- **"Nothing happens when executing DUE.jar"**

Ensure that the Java Runtime Environment (JRE) is installed on your machine and is in your PATH. The JRE should be version 5.0 (1.5.0) or higher. To check that a suitable version of the JRE is installed and in your PATH, open a command prompt and type:

java -version

If the command is not recognised, the JRE is not installed and in your PATH. If the version is below 5.0 (1.5.0) update the JRE (see above).

If this does not help, check the "C:\" directory for a log file named "due.log". If the first line of the log file is:

com/incors/plaf/alloy/AlloyLookAndFeel

then DUE has been unable to load the resources required for proper execution of the software. Check that "DUE.jar" has not been moved from the original installation directory (i.e. that the internal structure of the archive "DUE_3.1.zip" is preserved).

Otherwise, send the error message to the authors for advice on how to proceed (brown@science.uva.nl). If a "C:\" directory cannot be accessed on your machine, the log file will not be written. Contact the authors for advice on how to proceed.

- **"An error message is thrown when executing DUE.jar"**

If an error message is thrown by the JRE (i.e. a java error appears in the message), the error may be caused by the local installation of Java.

2.4 Altering memory settings

By default, the amount of RAM memory available to DUE is restricted by the Java Virtual Machine. In order to perform an uncertainty analysis with large datasets, it may be necessary to override this default and increase the amount of memory available. This is achieved by executing DUE on the command line (e.g. using a DOS prompt). Navigate to the installation directory of DUE, and type:

start javaw -jar -Xms**64**m -Xmx**500**m DUE.jar

where 64 is the minimum memory allocation in this example (MB) and 500 is the maximum allocation. The maximum memory allocation should be significantly lower than the total amount of RAM available, as other programs, including the operating system, will require memory to run without swapping (which slows everything down).

2.5 Source code and documentation

The Java source code for DUE can be found in the "src.zip" archive in the root directory of your installation. The Application Programming Interface (API) is described in the html documentation, which accompanies the software (/docs directory).

3. OVERVIEW OF FUNCTIONALITY

3.1 Summary of functionality in DUE Version 3.1

The functionality currently supported by DUE includes:

- The specification of a probability model for different types of attribute, including continuous numerical attributes (e.g. rainfall), discrete numerical attributes (e.g. bird counts) and categorical attributes (e.g. land-cover). The attributes may be constant in space and time or may vary in space or time. Combined space-time functionality is currently limited to spatial raster data (in 2D). Furthermore, an assumption of temporal independence is required when assessing uncertainty for spatial time-series (i.e. the uncertainties at different times are unrelated).

- The objects supported by DUE include spatial rasters, spatial vectors, time-series of rasters and simple time-series;

- The specification of a probability distribution function (pdf) for the positional uncertainty of 2D spatial vectors, including correlations within and between coordinates. Objects that comprise multiple points, such as lines and polygons, may be assumed "rigid" under uncertainty, where all internal coordinates move identically, or "deformable", whereby each internal point can move separately. The uncertainty of a rigid object is completely specified by a translation and/or rotation of that object about a single point. In contrast, the uncertainty of a "deformable" object requires the marginal uncertainties to be defined at all internal points, together with any relationships between them. For deformable objects that contain overlapping boundaries (duplicate points), such as field boundaries, the duplicate points may be grouped together, in order to maintain the boundaries when simulating from the pdf;

- Parametric pdfs for continuous numerical data (normal, lognormal, Weibull etc.) and discrete numerical data (Poisson, binomial etc.), with the option to define a non-parametric pdf, comprising user-defined outcomes and probabilities, for discrete numerical and categorical data;

- The use of expert judgement OR sample data to help define a probability model. Limited functionality is included for estimating a pdf with sample data, including estimation of pdf parameters and fitting a correlation model. In addition, samples are used to improve the accuracy of the simulated datasets by honouring these data during simulation (so-called 'conditional simulation'). Future releases of DUE will allow expert judgement and sample data to be combined within a Bayesian framework;

- The specification of correlations within a single attribute in space or time if the attribute values are normally distributed. These 'autocorrelations' are defined with a correlogram, whereby the correlation between two locations (two uncertainties) varies as a function of their separation distance, and possibly direction (2D/3D), but is otherwise constant in space and time. In this framework, the magnitude of uncertainty (variance) can vary at each point in space or time;

- The specification of correlations between attributes (crosscorrelations) if the attributes are continuous numerical and their pdfs are joint normally distributed. Cross-correlations are defined for pairwise relationships between attributes using correlation functions.

- Aggregation of (uncertain) attribute values to larger spatial or temporal scales, including aggregation from points to blocks, with the following restrictions:

  - Only continuously varying quantities, such as time-series and spatial rasters, can be aggregated (i.e. no spatial vectors);
  - The coarse scale must divide exactly by the fine scale in each coordinate dimension. In other words, a raster with 10m*10m cells can be aggregated to a raster with 50m*50m cells, but not one with 15m*15m cells;
  - For aggregation from one block (length, volume) to another block, the aggregation statistic must also commute between scales. A statistic commutes between scales if the aggregated value can be determined iteratively from groups of the input values (e.g. the mean commutes, but the median does not);
  - Aggregation from points to blocks is only supported for the mean statistic, as this can be estimated sensibly from small numbers of points;
  - Disaggregation is not supported.

- Simulation from pdfs for continuous numerical, discrete numerical and categorical attributes that vary in space or time for use in Monte Carlo studies with models. An exact, and fast, simulation routine is used for joint normally distributed pdfs if the correlation matrix is sufficiently small (or available memory is sufficiently large). Otherwise, simulation is conducted iteratively using the Sequential Simulation Algorithm. In most other cases, distribution-specific methods are used to simulate from the marginal pdfs.

- Simulation of positional uncertainties in 2D spatial vectors (as above).

- Import from, and export, to file with a limited range of formats, including ESRI Shape files for spatial vectors and ASCII raster for raster files. Saving an uncertainty analysis in a project file with a .due extension.

- Searching, retrieving and saving pdfs for time-series in a DUE-enabled Oracle-ArcSDE database.

- A simple Application Programming Interface for obtaining realisations of stored uncertainty models for use in external software (an alternative to file writing that requires a simple programmatic link between DUE and an external model).

3.2 Planned functionality

The upcoming functionality for Version 3.1 of DUE includes, in no particular order:

- Allowing UM to be defined for individual sources of uncertainty. In that case, the overall UM is the sum of models from each source of uncertainty;

- Incorporating statistical dependence within and between attributes that are not joint normally distributed. Initially, this will focus on autocorrelations in discrete numerical pdfs, such as the Poisson distribution, and in categorical attributes, for which Markov Random Fields appear promising. An ongoing challenge is to balance statistical realism with practicality in applying pdfs to environmental data.

- Extension of the library of sources of uncertainty, including links to external resources (online and offline);

- Extension of the range of uncertainty model structures to include confidence intervals and scenarios;

- Extension of the range of parametric pdfs, and inclusion of a non-parametric continuous pdf (non-parametric discrete pdfs are available);

- Extension of the online help functionality;

- Support for 3D raster data;

- Extension of the DUE-enabled database to store spatial rasters and vectors (currently limited to time-series);

- Integration of DUE within a data assimilation toolbox for recursive estimation of model states under uncertainty;

- Inclusion of methods for expert elicitation of probability models;

- Semi-automatic fitting of correlation functions.

4. GETTING STARTED

4.1 Performing an uncertainty analysis with DUE

Performing an uncertainty analysis with DUE is separated into five stages, namely:

1. Loading (and saving) data;
2. Identifying and describing the sources of uncertainty;
3. Defining an uncertainty model, aided by the description of sources;
4. Evaluating the "goodness" of the model;
5. Generating realisations of data for use in an uncertainty propagation analysis.

These stages are separated into 'panels' in the user interface. To begin with, an uncertainty analysis with DUE may involve linearly navigating through these panels using the "**Next**" and "**Back**" buttons. Such a linear navigation is useful when an uncertainty model has not yet been defined. After an uncertainty model has been defined and saved for an object or attribute of interest, the route of entry into the software may vary. For example, it might involve modifying and saving an existing model for later use or generating realisations of objects and their attributes for use in Monte Carlo studies.

On starting DUE, the first stage involves loading data from an existing project file or by starting a new project and loading data from file or database. Stages 2 (describing the sources of uncertainty) and 4 (evaluating the goodness of a model) may not be necessary, depending on the application of the software. Stage 2 is useful for structuring an uncertainty analysis by considering the major sources of uncertainty, including which sources cannot be included and how important they are in assessing uncertainty propagation (i.e. the 'propagation risk'). A skeleton library of uncertainty sources is provided and may be extended for this purpose. However, this functionality may be less useful if the sources are well-known and unambiguous. Similarly, assessing the goodness of an uncertainty model may not be necessary if the uncertainty analysis does not require detailed scrutiny by others.

4.2 Administrative functions

The opening window of DUE, together with the Taskbar, is shown in *figure 1*. The opening window displays the objects and attributes loaded into the software, together with details about their value scales and structures and whether an uncertainty model has been defined for them. The Taskbar is visible throughout the operation of DUE and is used for administrative tasks, such as creating, opening and saving a project, selecting objects and attributes, deleting them from a project, and loading data from a file or the DUE-enabled database. The Taskbar options are listed in *table 1*.

Shortcuts are provided on the Taskbar for some common operations, but all operations are otherwise accessible through the dropdown lists. After importing objects and attributes into DUE, one or more objects and their attributes may be selected in the opening window (*figure 1*) or via the drop down menus (one object/attribute only), which are visible throughout an uncertainty analysis (top right of *figure 1*). The "Input" and "Output" windows of DUE allow for the selection and simulation of any attributes currently loaded, respectively. All intermediate windows refer to the (uncertainty of the) single attribute selected in the Input window, as uncertainty models are constructed for individual attributes or iteratively, from individual attributes, in the case of joint models.

Figure 1: The opening window of DUE

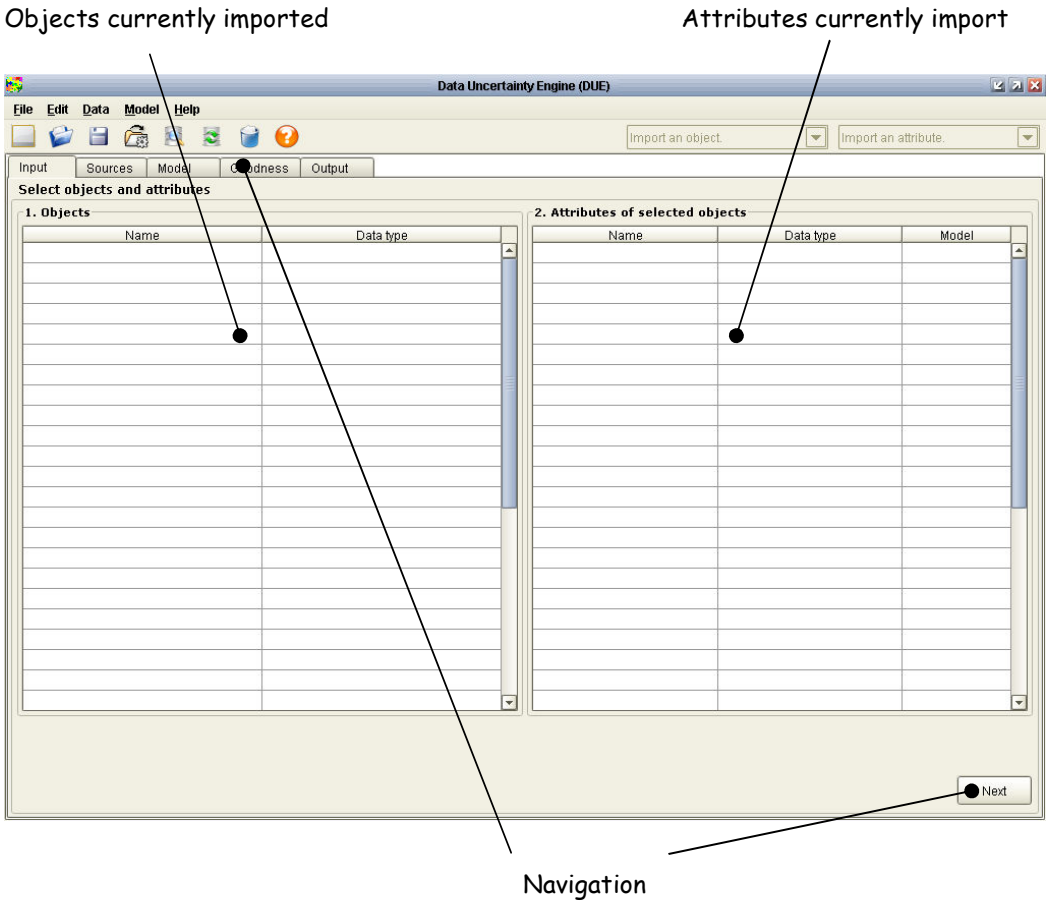


Table 1: Menu items

Menu	Function	Use
File	New project	Creates a new project
	Open project	Opens a project file (*.due)
	Save project	Updates or creates a project file (*.due)
	Save project as	Updates or creates a named project file (*.due)
	Link to external models	Stores a project to file for programmatic access
	Exit	Exits DUE
Edit	Remove item(s)	Remove selected objects/attributes
	Edit selected object	Edit the attributes of a selected object
	Edit null values	Enable null values for assigning uncertainties
Data	Load object(s) from file	Load objects and attributes from file
	Load object(s) from database	Load objects and attributes from a database
	Update object(s) in database	Update the uncertainty information in a database
	View scale information	Shows the scale of a selected attribute
	Add a constant object	Add an object that is constant in space and time
	Picture viewer (Disabled)	A data viewer for spatial and temporal objects
Model	Refresh selected model	Restores a saved uncertainty model
	Remove selected model(s)	Removes the selected uncertainty model(s)
	Probability options (Disabled)	Advanced options for probability modelling
	Correlation options (Disabled)	Advanced options for correlation modelling
Help	Messages on/off	Turns online help messages on/off
	Console	Shows the details of incorrect user actions
	About	Credits and conditions of use

4.3 Importing and exporting data with files

DUE supports uploading of information from file or from a database. In both cases, the information may be 'raw' (i.e. data for which an uncertainty model has not been defined) or data for which an uncertainty model exists. The latter includes a project file with the .due extension, where all information, including the uncertainty models and user interface settings for a project, are stored.

File formats:

File formats supported by DUE include:

- ESRI 'Shapefiles for spatial vector datasets (e.g. points, lines, polygons);

– A simplified GeoEAS file format for reading spatial point vectors with one or more attributes, and for writing realisations of spatial point vectors. An example of this format is given below:

```
spaceDim 2
"X" "Y" "Rainfall"
181072.0 333611.0 1022.0
181025.0 333558.0 1141.0
181165.0 333537.0 640.0
```

The first line of the header contains the `spaceDim` keyword, which refers to the number of spatial dimensions in the dataset, and may be set to 2 or 3. The second line contains the names of the attributes. In this case, the first two columns are interpreted as X and Y coordinates (`spaceDim 2`), regardless of the names provided. Other columns contain the attribute values, for which the attribute names ("Rainfall" in this case) are read from the header. The columns are separated by white space.

– ASCII Raster for 2D raster data (.asc). An example of this format is given below:

```
ncols         11
nrows         2
xllcorner     573000
yllcorner     181000
cellsize      10000
NODATA_value  -9999
0 7 7 18 18 18 7 7 7 7 18
0 7 7 18 18 18 7 7 7 7 18
```

The file header contains the number of columns in the raster grid (`ncols`), the number of rows (nrows), the lower left corner of the grid in arbitrary coordinates, including the X-coordinate (`xllcorner`) and the Y-coordinate (`yllcorner`), the size of the square grid cells (`cellsize`) and the value reserved for null or missing elements. The data values are separated by white space.

– An ASCII file for simple time-series (.tsd). An example of this format is given below:

```
Chloride Nitrogen
-9999.0 -9999.0
1990.01.11,65.0,9.6
1990.01.22,56.0,7.9
1990.02.06,44.0,11.6
```

The first line of the header contains the names of the attributes in the time-series (two attributes in this case). The second line contains the value reserved for null or missing elements in each attribute. The times are stored in the first column of the file as real dates in the format `yyyy.mm.dd.m.s.ms`. Integer-incremented dates are currently interpreted as years (this functionality will be extended). The differences between consecutive times may be regular or irregular. The additional columns contain the values for each attribute by name. The columns may be separated by white space or a comma.

Importing files:

The user-interface for importing data from file comprises two parts (*figure 2*), namely a Files Dialog (*figure 2a*) and an Objects Dialog (*figure 2b*). When importing objects and attributes from file, some of the information necessary to perform an uncertainty analysis may be missing. For example, some information about the scale or 'support' of the data may not be stored in file. In addition, it may not be possible to diagnose the attribute structure (e.g. continuous numerical) from the data structure (e.g. integer terrain heights). Both are important in performing an uncertainty analysis with DUE.

Figure 2a: the Files Dialog used to import data from file

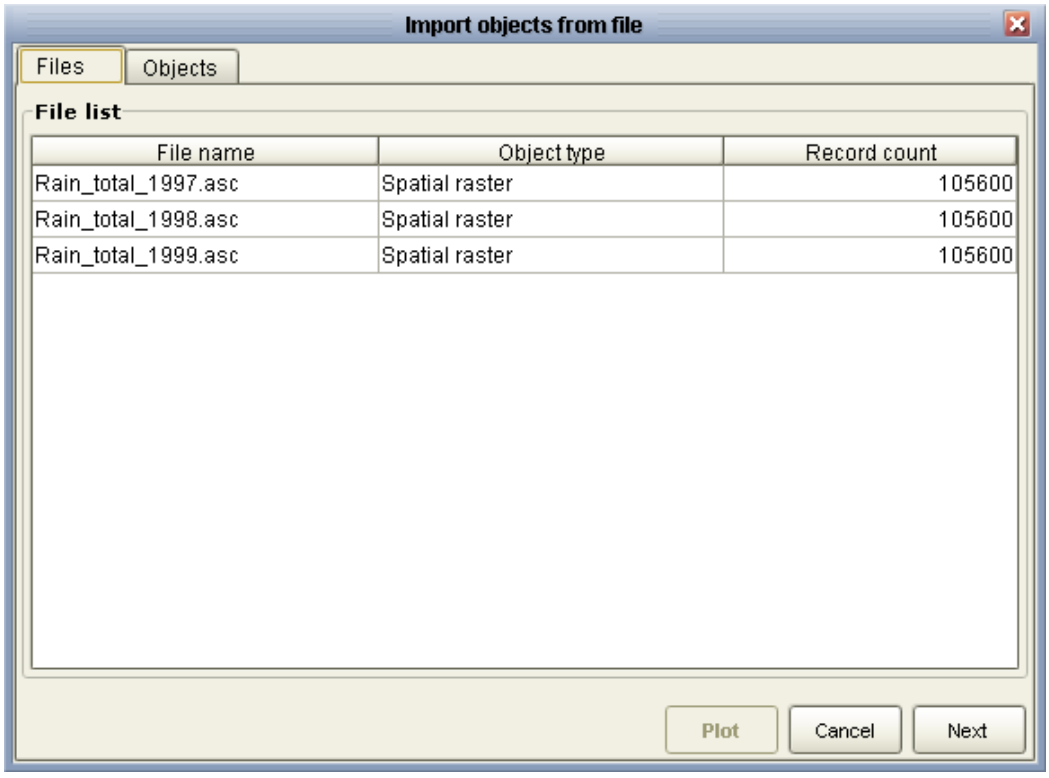
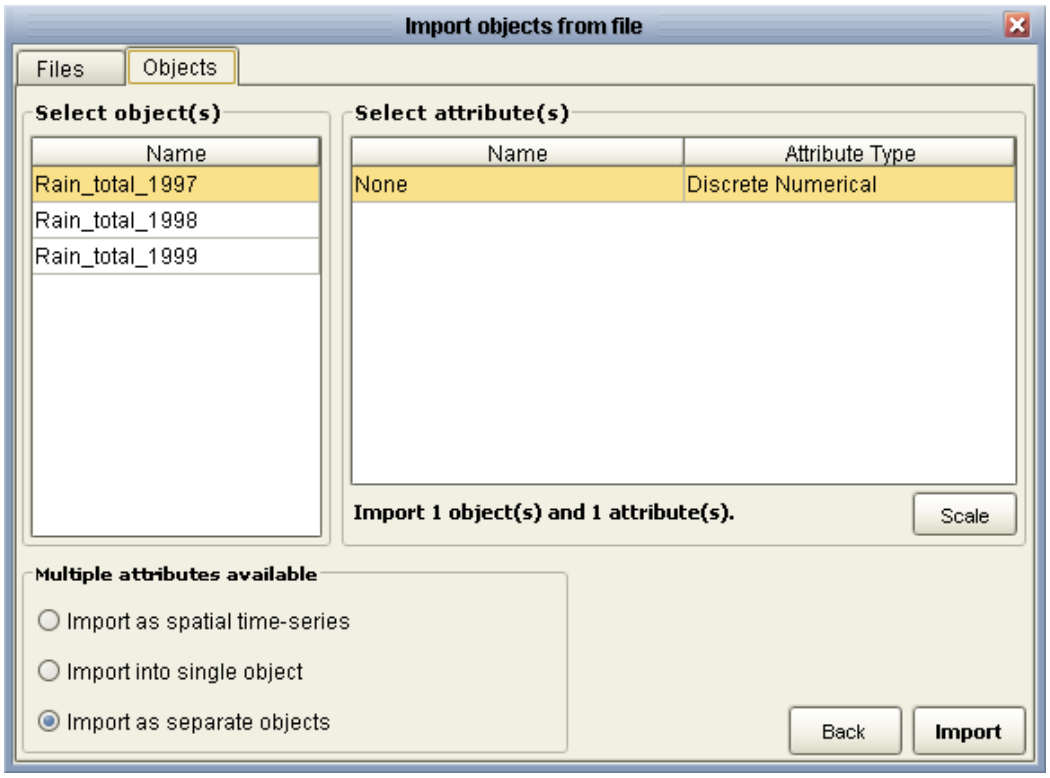


Figure 2b: the Objects Dialog used to import data from file



As indicated in *figure 2a*, the Files Dialog comprises a panel with information about the data read from file and a second panel requiring user input on how to construct an object from these data. This dialog is displayed after selecting one or more files to import. In future, the dialog will allow visualisation of datasets before importing them.

The Objects Dialog is revealed by clicking "**Next**" in the Files Dialog or selecting the "**Objects**" tab (*figure 2b*). The left table displays the names of the objects being imported from file, and the right table displays the names and data types of their associated attributes. When multiple attributes are imported at once, from one or multiple files, they may represent:

1) multiple attributes of a single object (e.g. different chemicals from one monitoring station)
2) single attributes of multiple objects (e.g. one chemical at multiple monitoring stations).
3) a time-series of one spatial attribute (e.g. landcover change)

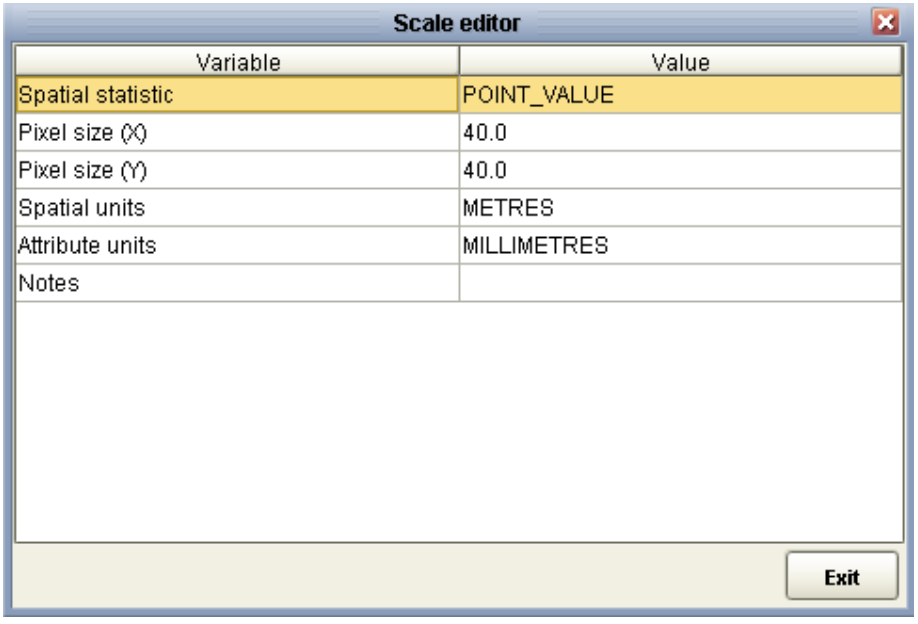
In practice, the difference between (2) and (3) are often semantic, but in DUE Objects are also used to collect attributes with equivalent supports. In other words, option (1) is only available if the support of the attributes is identical.

The attribute type is determined automatically from the data structure of continuous numerical attributes (decimal places) and categorical attributes (non-numeric data), but can be altered for discrete numerical data, as integers may refer to continuous attributes (e.g. rounded terrain heights) or categorical attributes (e.g. integer land-cover classes).

On importing attributes into DUE, the scale of the attributes must be defined. The information required will depend on the object/attribute type, but may include the period of aggregation or grid-cell sizes, the attribute units and spatial or temporal units. Where possible, this information is obtained from file, or from the database in which the attribute was stored. The scale dialog is accessed by selecting an attribute and clicking "**Scale**" (*figure 2c*).

Data are imported into DUE by selecting one or more objects/attributes and clicking "**Import**" in the Objects Dialog (*figure 2b*).

Figure 2c: the Scale Dialog used to import data from file

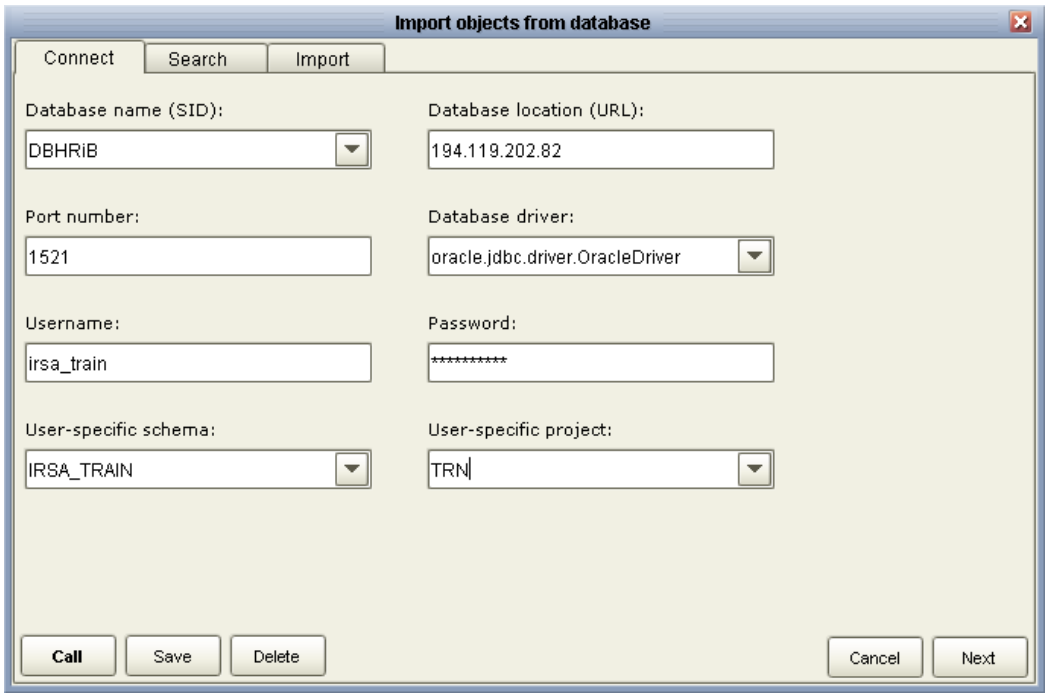


4.4 Importing and exporting data with a DUE-enabled database

A prototype database is available for storing, retrieving and editing uncertain objects and attributes with DUE. The database has been implemented in Oracle, with a link to ArcSDE for storing spatial data. Oracle and ArcSDE are proprietary software, but the database structure and administrative tools are freely available (contact brown@science.uva.nl). DUE cannot be connected to an arbitrary database, as it requires a specific structure for storing uncertainty information. In order to illustrate the database functionality, a remote server with the database software and a data library has been implemented for use with DUE. In order to use the DUE-enabled database, you will need to register your computer's IP address with the authors.

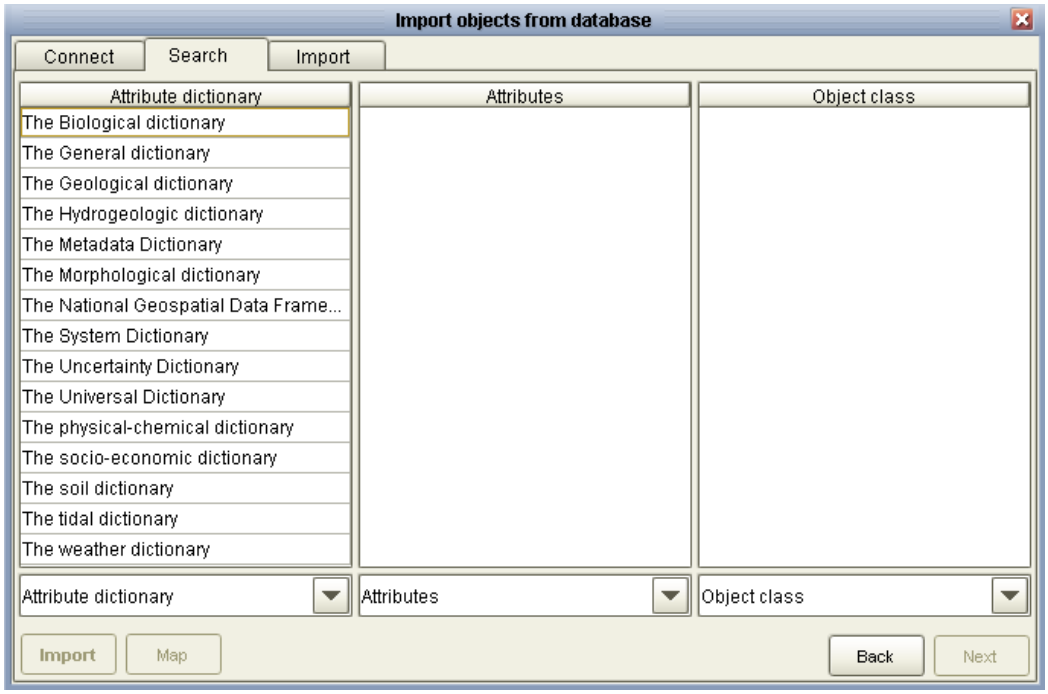
The user-interface for connecting to, searching and retrieving data from the database comprises three parts (*figure 3*), namely a connection dialog (*figure 3a*), a search dialog (*figure 3b*) and an import dialog (*figure 3c*). Once data have been retrieved from the database, the uncertainty information associated with those data may be added, removed or edited through **Data > update object(s) in database**, which has a Taskbar shortcut. Since no tools are provided for loading objects and attributes into a database (this software is available separately), the uncertainty information can only be updated if the objects and attributes were obtained from a database via DUE. In this case, the database parameters for each attribute are stored in a .due project file, which allows discontinuous updating of the uncertainty information.

Figure 3a: The user interface for connecting to a DUE-enabled database



The connection dialog (*figure 3a*) displays the parameters for connecting to a DUE-enabled database. These parameters include the name of the database or the Oracle system identifier (SID), the location or Universal Resource Locator (URL), the port number on the host server, the database driver and the username and password of a given user. The "**Call**" button is used to connect to the database, and results in the display of all schemas and projects available to a user. Once a schema and project has been selected, the "**Next**" button or "**Search**" tab can be used to display the Search Dialog (*figure 3b*). To view information in the prototype database, establish an Internet connection, enter `irsa_train` for the username, `irsa_train` for the password, and select the "`IRSA_TRAIN`" schema and the "`TRN`" project.

Figure 3b: The user interface for searching a DUE-enabled database



A combination of list selection and graphic visualisation are used to search the database for objects and attributes (only list selection is available in Version 3.1 of DUE). List selection employs a set of 'query models', representing routes into the database, to locate objects and attributes. The query models are located in drop-down menus at the bottom of the "Search" dialog. These menus also facilitate keyword searches on items in the tables (e.g. entering "soil" in first menu, followed by ENTER will filter the results by this keyword, displaying one item: "The soil dictionary").

The route into the database will depend on the types of objects and attributes required and the meta-information available to locate them, but multiple routes are

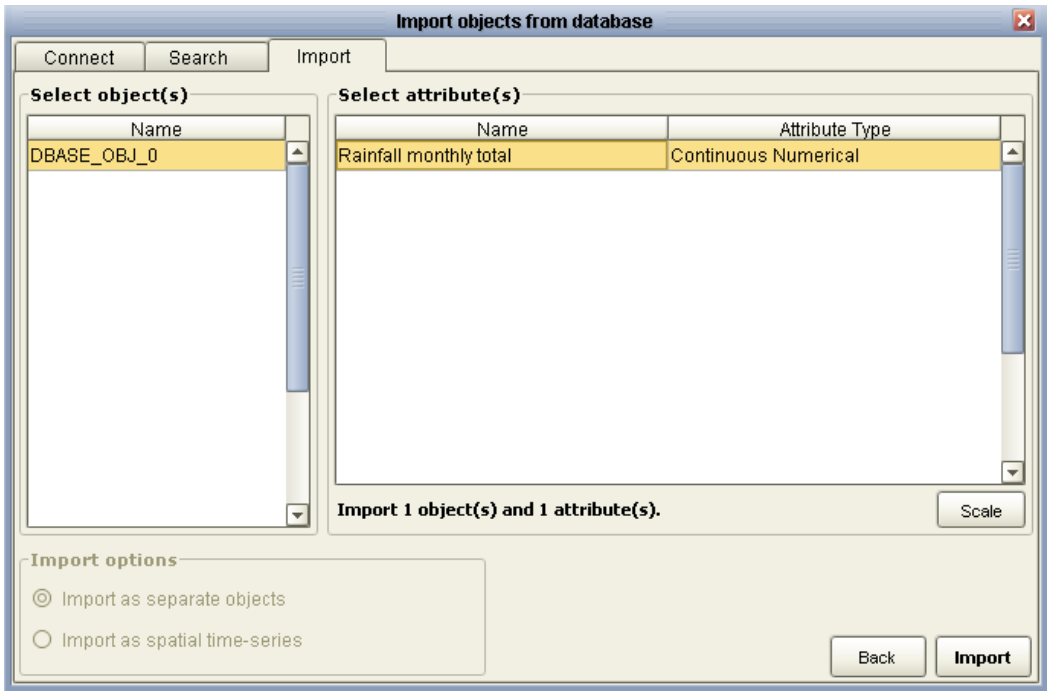
usually possible. The default search model begins with a list of "Attribute Dictionaries" used to collect similar attributes in the database. For example, "The weather dictionary" is used to locate meteorological attributes. In this model, the selection of an Attribute Dictionary leads to the display of all attributes associated with that dictionary. On selecting a particular type of attribute (e.g. Rainfall monthly total), the adjacent table reveals a list of all object classes at which that particular attribute is measured (e.g. object class Raingauge). The graphical viewer might then help to locate a specific object by displaying all objects, coded by class type, at which the attribute-type is measured (accessed via "Map", but not available in Version 3.1 of DUE).

If more detailed information is available about a *particular* object and attribute, the query model **Object class** > **Object** > **Attribute** can be used, and leads to the selection of one or more attributes at a specific object (e.g. a specific location) in three steps. Multiple objects or attributes can be imported at once. When one or more objects or attributes (or the criteria for locating multiple objects and attributes) are selected, the associated data can be imported with the "**Import**" button.

Detailed information about the conceptual structure and data tables used to store objects and attributes in a 'DUE-enabled' database, as well as the uncertainty information associated with them, can be found through www.harmonirib.com. In this context, it is sufficient to note that particular objects are identified in the database by their Object Identification Attributes (OIA). The OIA are set by the database user/maintainer at the point of loading objects. All such OIA are displayed in the user interface for particular objects.

The Import Dialog (*figure 3c*) displays further details on the objects and attributes selected from the Search Dialog for import into DUE, including the object/attribute names, the attribute data type, and any scale information associated with it (accessed via the "**Scale**" button after selecting an attribute). The objects and attributes can be renamed here. The "**Import**" button is used to import the data into DUE.

Figure 3c: The user interface for importing objects and attributes from a database



4.5 Creating projects

All work within DUE (including user interface settings) can be saved to a project file with the .due extension once an object has been loaded from file or database. A project is saved using the **Save** or **Save As**… option in the **File** dialog, or the shortcut to **Save** on the Taskbar. Project files are stored in a binary format and are not, therefore, human readable or editable. An XML version of the project file will be available in a future release of DUE.

5. EXAMPLES AND EXERCISES

The basic functionality of DUE is illustrated in the following examples and exercises. The exercises should be conducted in sequence, as each builds on the expertise gained in the previous ones. The assumptions made in the examples are purely illustrative, and are not necessarily realistic for other applications of DUE.

5.1 Importing a time-series into DUE from file

Go to the opening window of DUE (*figure 1*). Execute **Data > Import object(s) from file**. A file chooser will appear. Navigate to the "due/resources/exampledata" folder in the root directory of your installation (e.g. C:/Program Files/DUE_3.1/due/ resources/exampledata) and open the file named "Chloride_Nitrogen.tsd". The Files Dialog will appear (*figure 2a*).

Click "**Next**" to enter the Objects Dialog. You can rename the objects and attributes by double-clicking on the relevant table cells. The data structure of both attributes is continuous numerical (decimal places were found) and cannot be altered. In order to import the attributes into DUE, some scale information must be defined. Click on the "**Scale**" button to enter this information for the "Chloride" attribute. The time series includes chloride samples that were measured instantaneously, so the temporal statistic POINT_VALUE should be selected. As the time-series includes actual dates, the time units are unambiguous. Nevertheless, you must specify a preferred time unit, as this will be the standard unit for working with these data in DUE (e.g. when defining a correlation function). Select "MONTH" as the temporal unit and "MICROGRAM/LITRE" as the attribute unit (type e.g. "MIC" in the attribute unit box to reduce the drop-down list of options). Text may be entered into the drop-down menu for attribute units, in which case the units are completed automatically when a unique match is found. Click "**Close**" to exit the dialog. Select only the Chloride attribute and click "**Import**" in the Objects Dialog (*figure 2b*) to import the attribute into DUE.

5.2 Importing a time-series into DUE from the prototype database

Register your computer's IP address with the authors (brown@science.uva.nl). Establish an Internet connection. Go to the opening window of DUE (*figure 1*) and execute **Data > Import object(s) from database**. The "Connect" dialog will appear. Enter `irsa_train` for the username and `irsa_train` for the password, then "**Call**" to attempt a connection with the remote database. If the connection is made successfully, a "Connected" message will be displayed and the "Schema" menu (*figure 3a*) will be updated with the database schemas available, otherwise an error message will be displayed. Select the "`IRSA_TRAIN`" schema, the "`TRN`" project and

click “**Next**” to enter the Search Dialog (*figure 3b*). The first table will be updated with the Attribute Dictionaries available in the database (see *Section 4.4* also).

The aim here is to import the `Rainfall monthly total` attribute of a raingauge in Greece. The raingauge is identified by its Object Class, `RNGS`, its Country Code, `GR`, and its Site Code, `AGBAR_001`. Given this information, the `Rainfall monthly total` attribute can be found in several ways. The routes for finding information are listed in the drop-down menus at the bottom of the Search Dialog (see *Section 4.4*). For example, you can search by `Attribute dictionary`, selecting `The weather dictionary`, then by `Attribute`, selecting `Rainfall monthly total`, then by `Object`, which lists all objects in the database where `Rainfall monthly total` is measured.

Select `Object class` in the first drop-down menu. Use the same menu box to search for the object class `Raingauge` in the list of results: delete the text `Object class` and enter `Rain` (note case sensitivity), then press ENTER. A single result, `Raingauge`, is displayed. Select `Raingauge` to populate the next table with all objects in the database from the Object Class `Raingauge` or `RNGS`. Search for the relevant object using the identification attributes (Object Class, `RNGS`, Country Code, `GR`, and Site Code, `AGBAR_001`) and select this object. Note that the bars separating each table can be moved to aid visualisation. The attributes of this object will be displayed in the final table. Select the `Rainfall monthly total` attribute and click “**Import**“ to import these data. Wait for the data to download from the host server (should be < 5 minutes). On successful download of the rainfall attribute, click “**Next**” to enter the Import Dialog (*figure 3c*).

The “Import” dialog displays the default names of the object/attribute, the attribute data type (Continuous Numerical), and the scale information associated with it (accessed via the “**Scale**” button, which is enabled on selecting an attribute). In the Import Dialog, click “**Import**“ to load the object into DUE. Notice in the opening window of DUE (*figure 1*) that an uncertainty model of type PDF (Probability Distribution Function) has already been defined for this attribute. You may select the rainfall attribute and navigate through DUE using the “**Next**” and “**Back**” buttons for a preview of how a pdf is defined in DUE (note that no information on the Sources of uncertainty or Goodness appears in this example).

5.3 Defining an uncertainty model for a time-series

Using the data imported in *Section 5.1*, the aim of this exercise is to define a simple uncertainty model for a time-series of chloride measurements. The time-series should appear in the opening window of DUE (*figure 1*), where some information

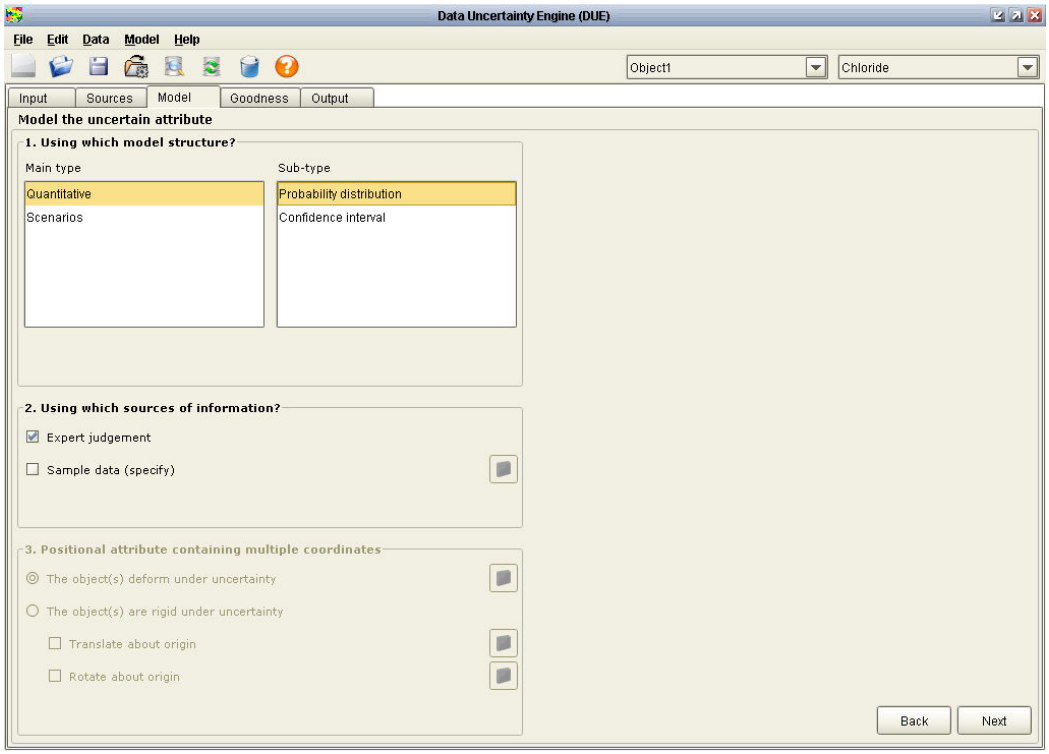
about the imported object (left table) and attribute (right table) is displayed. Before going further, you can now save a project using the **Save** or **Save As**…. options in the "File" menu. Save the project and re-open DUE. Open the saved project and the newly imported time-series object should re-appear.

Notice that the uncertainty of the attribute has not yet been defined (right table). When multiple objects and attributes are imported into DUE, the selected object(s) and attribute(s) are 'active', and the subsequent windows will be updated according to the selection made.

The "Sources" dialog will not be used here. Ensure the time-series is selected, and then navigate to the "Model" dialog by clicking "**Next**" twice or by selecting the "Model" tab.

In the first window of "Model" (*figure 4*), an uncertainty model structure is chosen for the active object and attribute. Only probability models are available in Version 3.1 of DUE. In future, confidence intervals and scenarios will be added, as they are more appropriate when information on uncertainty is limited. **Select Quantitative > Probability distribution** in the first "Model" window.

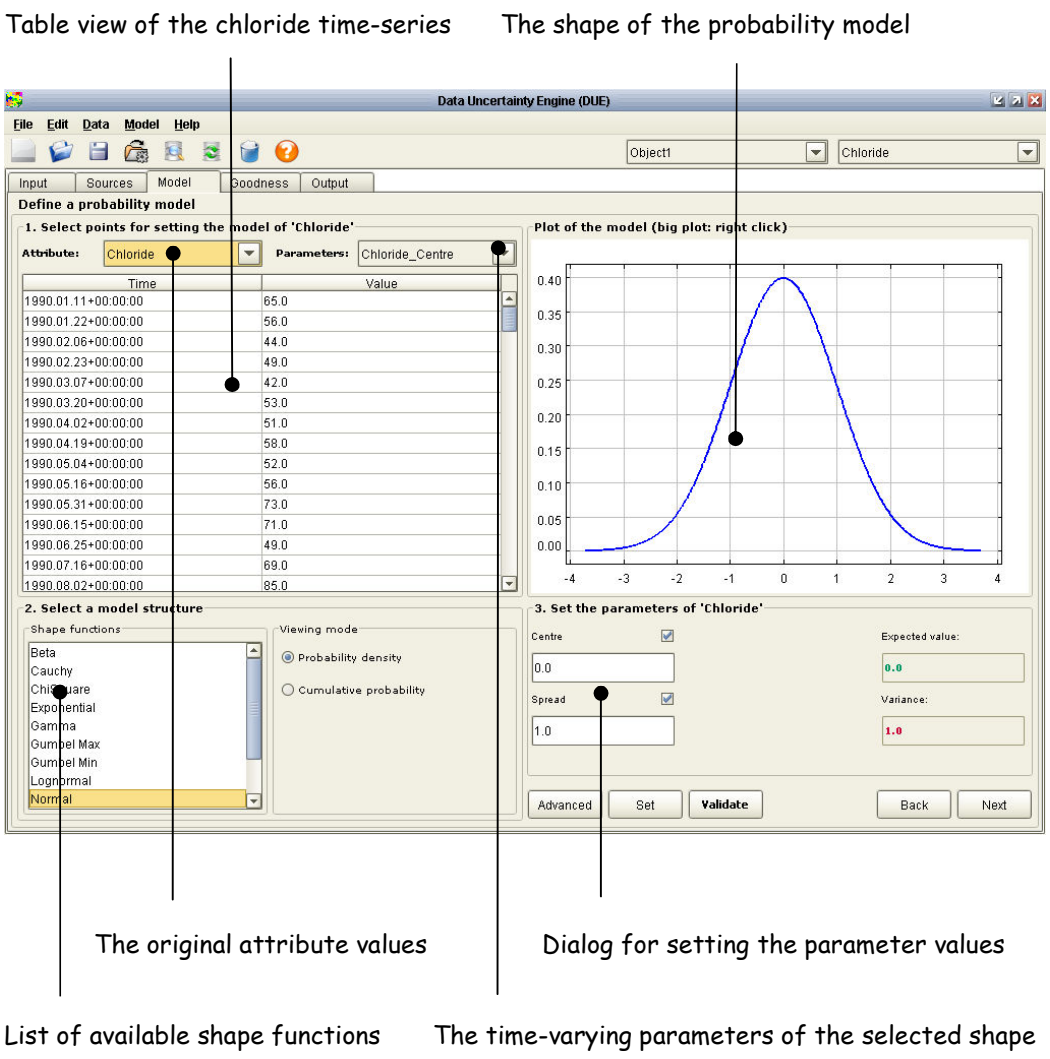
Figure 4: the first model dialog for selecting an uncertainty model structure



Two options now appear for quantifying uncertainty with a probability model. The first option allows ONE of two sources of information to be selected as the basis for assessing uncertainty, namely "**Expert judgement**" and "**Sample data**". In future, a Bayesian combination of these two information sources will be allowed (i.e. a prior based on expert judgement and a posteriori updated with sample data). Samples have two purposes in DUE, namely: 1) to help estimate the parameters of an uncertainty model; and 2) to improve the accuracy of the realisations locally by honouring the (certain) sample data. Sample data will not be used in this exercise (see Exercise 5.6); select "**Expert judgement**" instead.

The second option refers to the positional uncertainty of objects that comprise multiple points and is activated by the selection of a positional attribute in the "Input" window (see Exercise 5.7). Click "**Next**" to display the next window (*figure 5*):

Figure 5: Assigning a probability model for each point in a time series



The second Model window (*figure 5*) is used to define a probability model for each point in the chloride time-series. Notice that the time series contains some NULL values caused by instrument failure on these dates. By default, null values are ignored when defining a probability model, but may be edited by selecting **Edit > Edit null values**.

In order to define a probability model for each point in the time series, a simple shape function and its parameters values must be defined at each location shown in the table. A shape function is selected using the scrollable list in the bottom left corner of the dialog. Only ONE shape function can be selected for all locations/times in the dataset, but the parameter values can vary at each location/time. Select a "Normal" distribution. Notice that the drop down box marked "Parameters", and the text boxes for setting the parameter values, have changed to match the selected distribution ("Centre" or mean and "Spread" or standard deviation for the normal distribution). Select the "Centre" parameter in the drop down box of parameters. The values in the table all change to '?', indicating that the parameter has not yet been set. The dataset (attribute or parameter) currently displayed in the table is highlighted orange in the drop-down menu. Once parameter values have been entered and validated, the model cannot be altered until the existing parameter values have been deleted (a prompt will appear).

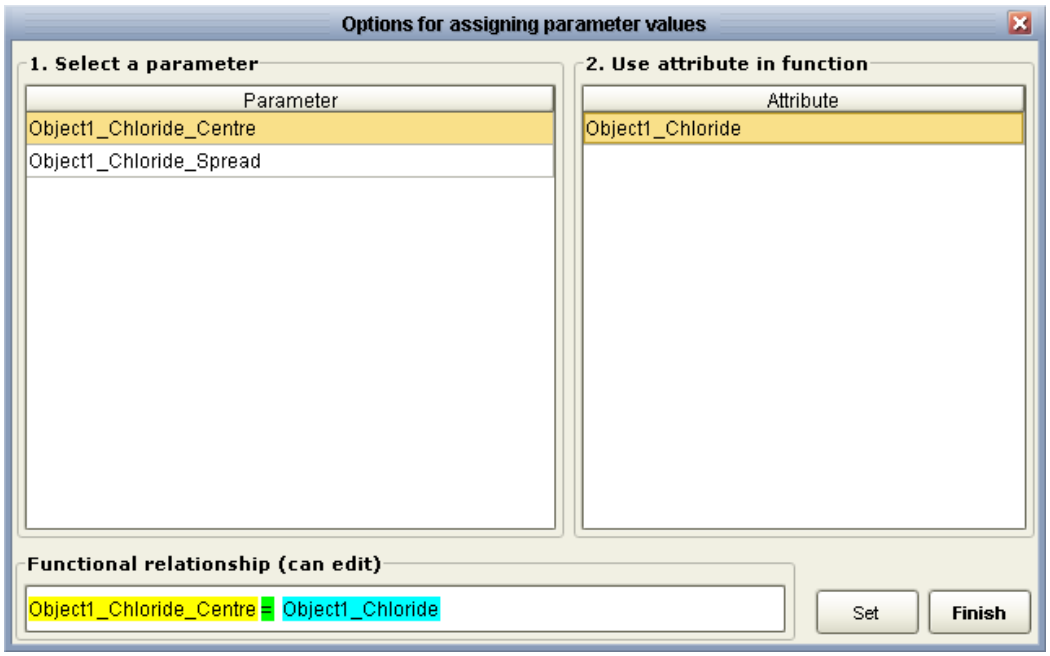
In the absence of sample data (i.e. expert judgement only), parameter values can be set in one of two ways, namely:

1. By selecting locations in the table, entering values in the parameter text boxes (i.e. **'3. Set the parameters'**) and clicking "**Set**" (if no cells are selected, the parameters are assigned globally). To select all locations at once, right click with the mouse and choose "**Select all points**". To select specific attribute values based on logical search criteria (<, > etc.), right click with the mouse and choose "**Custom selection**".

2. By selecting "**Advanced**" and setting the parameter values using existing attribute values (*figure 6*). In this case, the parameter values can be set as a function of the attribute values, or simply as the attribute values themselves. For example, the centre parameter of the normal distribution might be assumed equal to the original data values and the spread (uncertainty) may be 10% of the original data values.

Click on "**Advanced**" to define the parameter values in this case (*figure 6*).

Figure 6: Options dialog for setting the parameter values of a pdf



Attribute values are assigned to a model parameter by selecting the relevant parameter (left table) and attribute (right table) and clicking "**Set**". Optionally, the functional relationship between the attribute and parameter can be edited. A wide range of functions, including arithmetic operators, is supported. Recognised functions are highlighted: model parameters receive a yellow highlight, attributes a blue highlight, mathematical operators a green highlight and numerical constants a red highlight. Operator precedence (*, then /, then - and + etc.) may be overridden using brackets.

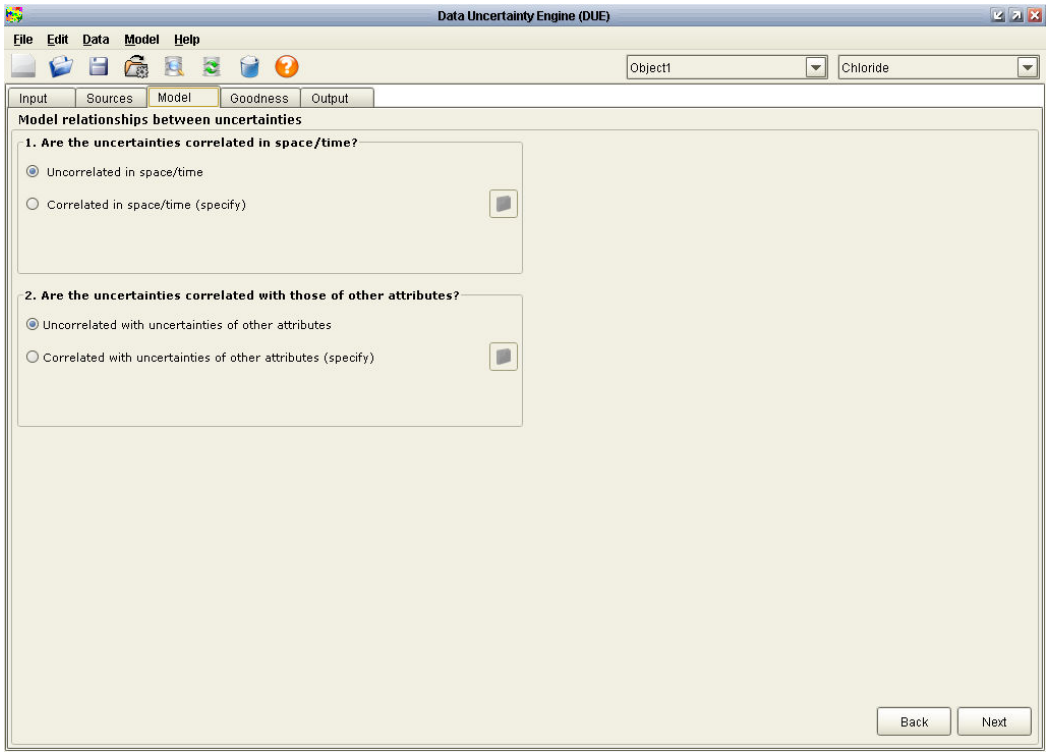
Assign the "Object1_Chloride" attribute to the "Centre" parameter of the normal distribution, and set the "Spread" parameter to 10% of the "Object1_Chloride" attribute by changing the functional relationship to: Object1_Chloride_Spread = Object1_Chloride * 0.1, assuming the object and attribute was not renamed on import. Click "**Set**" to assign the parameters and then "**Exit**" to return to the "Model" dialog. Check the new parameter values by selecting a parameter in the drop down box above the table (*figure 5*).

To validate the parameter values and save the model, click "**Validate**". A probability model has now been defined for each point in the Chloride time-series. On selecting a point in the table, the values shown in the parameter text boxes, together with the graphical display of the shape function, correspond to the marginal distribution of the selected time.

5.4 Defining a correlation model for an uncertain time-series

Using the chloride time series from the previous exercises, a correlation model will now be defined for the uncertain time series. In the presence of correlation, persistent lengths or 'patterns' will appear in the realisations of the time-series. In the absence of correlations, random patterns will appear in the realisations. Correlations may occur if the measurement errors vary with sampling conditions. For example, overestimation may occur in some conditions and underestimation in other conditions. These correlations will influence the simulated output by leading to systematic changes in chloride values at adjacent times. Indeed, in this example, the correlations will depend *only* on the separation distance (period) between measurement times. This assumption is not necessary, but greatly simplifies the estimation of a correlation matrix, which would otherwise need to be specified (i.e. ((296 * 296)/2)-296 = 43,512 correlation coefficients for this small dataset), and is often a reasonable assumption. If it is realistic, the correlation coefficients can be determined from a simple function or 'correlogram' comprising only one parameter in the simplest case, namely the 'average correlation length' or the distance at which the attribute values are no longer correlated (depending on the function chosen).

Figure 7: Defining relationships between uncertainties

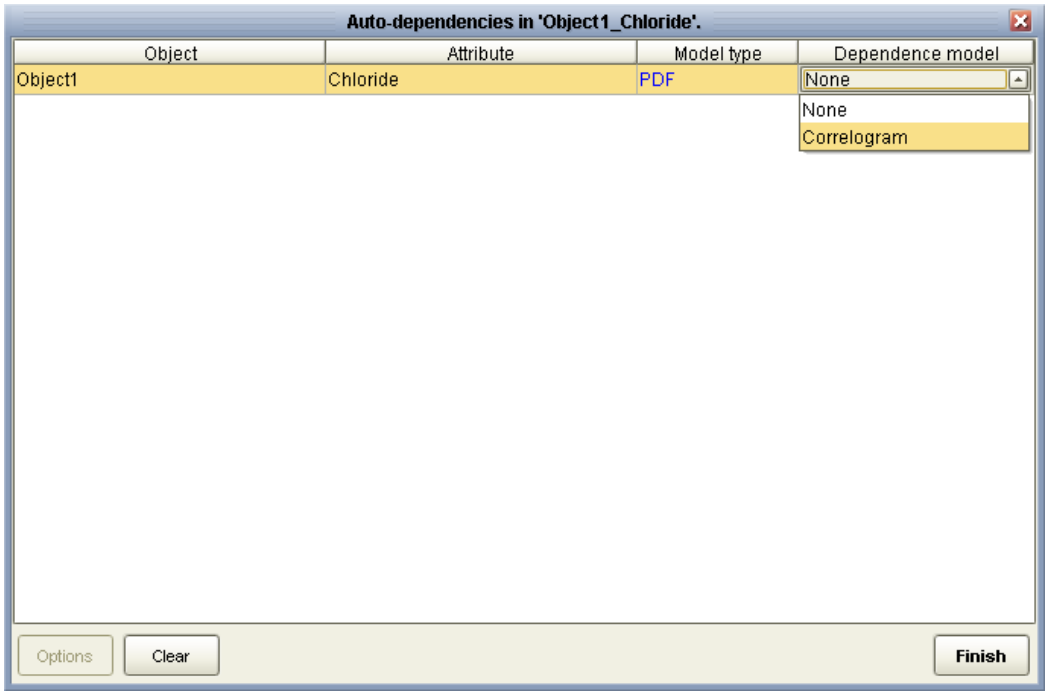

Data Uncertainty Engine (DUE)
File Edit Data Model Help
Object1
Chloride
Input Sources Model Goodness Output
Model relationships between uncertainties
1. Are the uncertainties correlated in space/time?
Uncorrelated in space/time
Correlated in space/time (specify)
2. Are the uncertainties correlated with those of other attributes?
Uncorrelated with uncertainties of other attributes
Correlated with uncertainties of other attributes (specify)
Back
Next

After completing Exercise 5.3 (above), navigate to the first correlation window (*figure 7*) by clicking "**Next**" from the window for marginal pdfs (*figure 5*). The window

comprises two options, the first for defining correlations within the selected attribute (autocorrelations) and the second for the defining pairwise correlations with other attributes in the “Input” window (cross-correlations). Currently, correlations can only be defined for attributes whose uncertainties are joint normally distributed (i.e. for cross-correlations a normal pdf must have been defined for each attribute before the dialog is accessible). Correlations assume a linear relationship between the marginal uncertainties; other forms of statistical dependence are not supported in DUE.

Select “**Correlated in space/time**” and click the newly enabled folder icon to define an autocorrelation model for the Chloride time-series. The resulting dialog (*figure 8*) shows the model structures available to specify the correlations between times; only correlation functions are available at present. Select the “**None**” option under the “**Dependence model**” column of the table and change to “**Correlogram**”. Click “**OK**” to exit and return to the main dialog. Click “**Next**” to open the window for defining correlation functions.

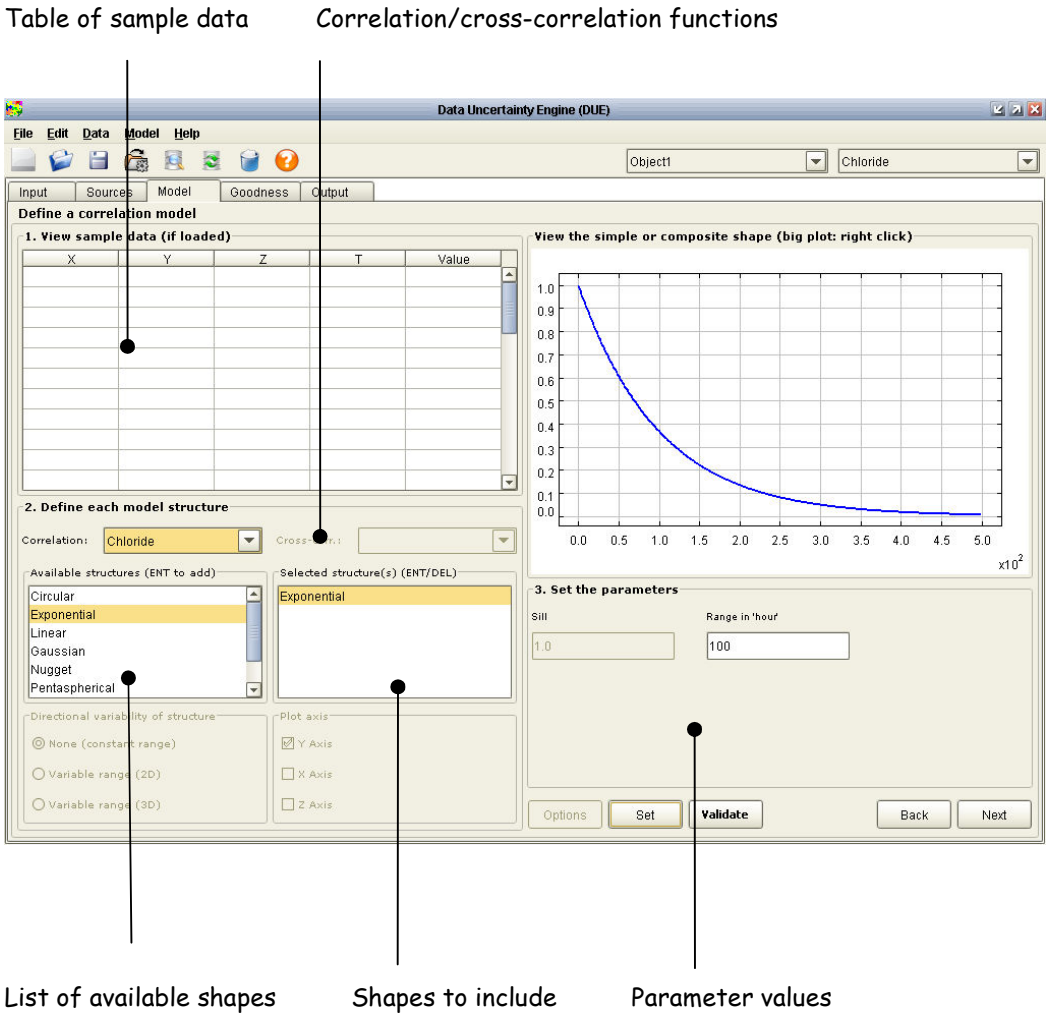
Figure 8: Selecting a model for dependencies between uncertainties



The window for defining correlograms is shown in *figure 9* and comprises a table for viewing sample data (filled if sample data were selected in the first model window), a drop-down menu for selecting autocorrelation functions (spatial or temporal correlations within a single attribute) and a menu for selecting cross-correlation functions (correlations between the uncertainties of multiple attributes). It also includes a list of shapes for building a correlation function, and a dialog for entering

the parameter values of each shape. For two- and three-dimensional attributes, the correlations may vary with direction as well as separation distance, for which further options are provided. You may right click on the plot to show a larger picture of the correlation function.

Figure 9: Defining a correlation function via expert judgement



In this example, the correlation function will be defined from expert judgement alone, as sample data are not available. Select an Exponential shape from the list of available shapes and press ENTER. Since the model comprises only one shape, the maximum correlation coefficient (1.0) is divided one way i.e. the "Sill" parameter is 1.0. Thus, the only parameter required is the average correlation length or 'range'. Note that the range is a *scaling parameter*, rather than the point of zero correlation, depending on the specific shape chosen (e.g. as in the exponential, but not the circular). You can experiment to view the impact of selecting different ranges on the simulated output. For now, set the correlation length to 100 months and Click "**Set**".

Click "**Validate**" to store the correlation model. On clicking "**Validate**", an attempt is made to create and factorise a correlation matrix for the selected attribute. If the matrix is too large, it will not be created and a slower algorithm (the Sequential Simulation Algorithm) will be used to generate realisations of the uncertain attribute. The probability model is now ready for simulation (see below).

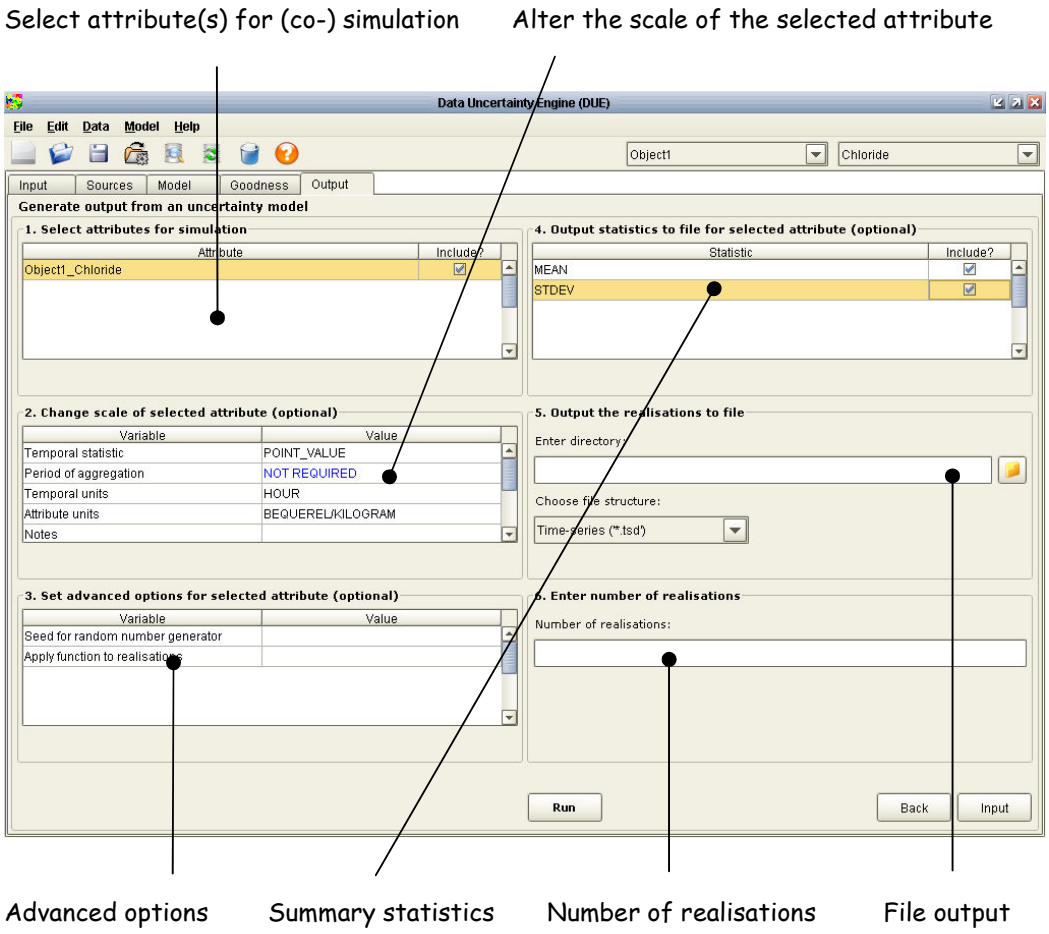
5.5 Generating realisations of an uncertain time-series

Using the probability model from *Section 5.3* or *5.4* it is now possible to generate realisations of an uncertain time-series.

Navigate to the "Output" dialog (*figure 10*), ignoring the "Goodness" dialog. The "Output" dialog provides various options for generating realisations of uncertain attributes for use in Monte Carlo studies with models. In order to simulate from an uncertainty model, the number of realisations and location for writing data (currently only files) must be provided for each uncertain attribute. Advanced simulation options are also provided, which vary with the selected attribute (e.g. for sampling with the Sequential Simulation Algorithm when a correlation matrix is not available). In addition, but under the restrictions listed in *Section 3*, the output scale of the realisations may be increased (i.e. aggregated). In simulating from a probability model, the realisations must honour the marginal probabilities at each location/time in the dataset, as well as the correlations between points. This can be checked by writing summary statistics for the realisations. For example, the mean and standard deviation should correspond to the parameter values shown in the second "Model" dialog (for the normal distribution). However, as these statistics are computed from sample data, the quality of the match will depend on the number of realisations created, increasingly linearly with that number.

Activate the Chloride attribute in the "**1. Select attributes for simulation**" table. Select the MEAN and STDEV for inspection and enter a directory for storing the output (either manually or by selecting a file with the adjacent button). Finally, enter a number of realisations to return (e.g. 100). Only one file type is available for writing time-series data, namely the .tsd type. Click "**Run**" to generate the realisations.

Figure 10: Simulating from a probability model

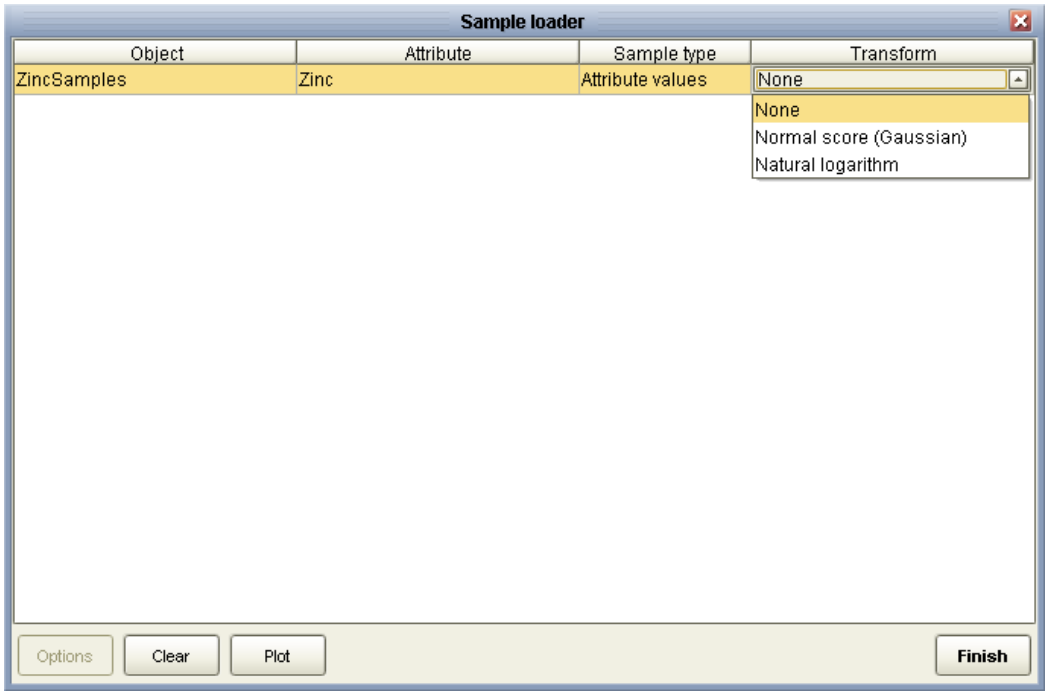


5.6 Generating realisations of a spatial raster attribute with sample data

Using the file chooser, navigate to the “due/resources/exampledata” folder in the root directory of your installation (e.g. C:/Program Files/DUE_3.1/due/resources/exampledata) and open the file “Zinc_base.asc”. The file contains a grid of empty values for which estimates of Zinc are required. Open “Zinc_base.asc” and import the object with a “Spatial statistic” of “POINT_VALUE”, “Spatial units” of “METRES”, and “Attribute units” of “MICROGRAM/KILOGRAM”. In this example, a limited set of observations are available to estimate the Zinc concentrations at unsampled points. For simplicity, it is assumed that the gridded predictions of Zinc are required at the ‘point support’ (cell centre positions), although a change of support (e.g. point to block) is also possible in DUE (for this, the spatial statistic should be set to MEAN). The observations of Zinc are located in the “Zinc_obs.eas” file of the “Example_data” directory. Open the “Zinc_obs.eas” file and import the observations into DUE with the same scale information as “Zinc_base.asc” (you will need to rename the object if the default object name was retained for “Zinc_base.asc”).

Navigate to the “Input” dialog and select the uncertain base map (originally “Zinc_base.asc”) for which predictions of Zinc are required. Move to the first “Model” dialog and select “Quantitative probability” and “Probability distribution” as the model type. As samples are available to help define the uncertain Zinc concentrations, they should be defined here. Select “**Sample data (specify)**” and click the newly enabled folder icon. This opens a sample loader (*figure 11*) comprising a list of objects that are recognised by DUE as ‘sample data’ (in this case, 2D points with the same numerical scale as the Zinc base map).

Figure 11: Sample loader used to view and select sample data in DUE



Sample data have two uses in DUE, namely: 1) to help estimate the parameters of a probability model, including those of a pdf and autocorrelation function; and 2) to improve the local accuracy of Monte Carlo realisations, by honouring the (possibly uncertain) sample points, as well as the overall probability model, during simulation. In this way, sample data are combined with a model of the underlying process to estimate an uncertain attribute. Linear regression (Kriging) is used to estimate attribute values at unsampled locations. Clearly, expert judgement is important here, as the properties of the sample data will rarely correspond exactly to those required by the probability model. For example, in using observations to improve the local accuracy of Monte Carlo realisations with DUE, the underlying process must be assumed joint normally distributed. Furthermore, the sample data should be approximately normally distributed to avoid unrealistically high prediction variances (uncertainties) in the simulated output. A common approach in spatial statistics is to

assume joint normality of the underlying process, and to transform the observations to their normal score values (i.e. a normal distribution) before conducting (2). The realisations are then made for normal-scores, and must be back-transformed to the original value scale after simulation. This is not straightforward, however, because many of the simulated values will not have a matching sample in the original observations, for which interpolation or extrapolation (i.e. transformation beyond the range of the sample data) is required (in DUE this involves linear interpolation within the range of observations and a power model, which may be altered, for the tails).

Select the observations in the sample loader (*figure 11*) and click "**Plot**" to view a histogram of the untransformed data values. The "Transform" column in the sample loader is used to transform the original data values (currently aimed at the normal distribution). Select the "Normal score (Gaussian)" transform and click "**Plot**" to display the normal score values of the sample data. Click "**OK**" in the sample loader to accept the normal score transform. In this example, the observations are 'attribute values', as they refer to direct measurements of Zinc. In other cases, the samples may refer to 'error values' (e.g. the difference between a remotely-sensed map and point observations) rather than attribute values, in which case the "errors" would be simulated and subtracted from a user-specified mean Zinc concentration.

Click "**Next**" to open the second "Model" window. Since a normal score transform was applied to the observations, the Zinc attribute is assumed joint normally distributed. In this framework, the mean of the sample data is taken as the "Centre" parameter of each marginal distribution (first-order stationarity) and the standard deviation of the samples is assigned to the "Spread" parameter (second-order stationarity). These initial estimates can be modified (e.g. with expert knowledge), but the "Centre" and "Spread" parameters must remain stationary. Click "**Validate**" to validate and save the probability models for each location and "**Next**" to enter the first correlation dialog. Select a correlogram model for the uncertain zinc values by activating the "Correlated in space/time" option, clicking the newly enabled folder icon, and choosing "Correlogram." Click "**OK**" to return to the main dialog and "**Next**" to enter the correlogram window (*figure 9*).

The graph window in the "Correlation" dialog shows the correlation between samples as a function of their separation distance or 'lag' in fixed intervals (similar to a histogram), while the table shows the transformed sample values. The assumption of stationarity is continued here, as the calculation of a correlation coefficient requires multiple samples of the same process, which are only available if the process is assumed constant in space. Automatic fitting of a correlation function to sample data is not available in DUE Version 3.1. Instead, the function must be fitted visually, or the parameters optimised with an external tool. In this example, an "Exponential"

shape function with a range of 400m fits the samples adequately. Assign this model and click "**Validate**" to validate and save the model.

Navigate to the "Output" dialog. Select the Zinc attribute and enter a number of realisations to return (e.g. 500). Aside from the number of realisations required, the simulation time will depend on the number of observations included in the (local) regression of Zinc, which may be reduced at the expense of local accuracy. Often, distant observations will have little influence on the simulated value at any given point (their contribution is weighed by the correlation function), but significantly increasing the computational load. In "**3. Set advanced options for selected attribute**" enter 30 as the maximum number of samples to include in the Kriging window. Select the MEAN and STDEV (standard deviation) for inspection and enter a directory for storing the output (either manually or by selecting a file with the adjacent button). Only one file type is available for writing spatial rasters, namely the ASCII Raster (.asc) type. Click "**Run**" to generate the realisations. Notice that the observations are honoured in each realisation and that the standard deviation (average uncertainty) of the predictions declines around the sample points.

5.7 Generating realisations of spatial vector objects

Using the file chooser, navigate to the "due/resources/exampledata" folder in the root directory of your installation (e.g. C:/Program Files/DUE_3.1/due/resources/ exampledata) and open the file "build.shp". The file contains a series of 2D polygons representing building outlines. Open the file and navigate to the "Objects" dialog (*figure 2b*). In order to import an object into DUE, at least one attribute must be specified. Select the "AREA" attribute and open the scale editor (*figure 2c*). Enter "POINT_VALUE", "METRES" and "METRE/SQUARED" for the "Spatial statistic", "Spatial units" and "Attribute units", respectively. Import the object into DUE. A positional attribute, comprising two coordinate vectors (X and Y), is added automatically using the "Spatial units" defined for the first attribute.

Select the "POSITION" attribute in the attributes table (*figure 1*) and navigate to the first "Model" window (*figure 4*). In DUE, objects are classified according to the movements allowed under positional uncertainty, and include: 1) rigid objects, where all points move with a constant relative motion; and 2) deformable objects, where each point can move separately from the surrounding points (see Appendix A for details). In this framework, the positional uncertainty of a rigid object is completely represented by the uncertainty of a single point, or origin, to which all other points are referenced. In contrast, the positional uncertainty of a deformable object requires an uncertainty model for every point associated with that object.

Select "Quantitative probability" and "Probability distribution". Assume that the "object(s) are rigid under uncertainty" and that the uncertainties involve a simple

translation of each object ("Translate about origin"). In DUE Version 3.1, the translation and rotation of a rigid object is made about the centroid of that object. In future, it will be possible to specify a custom origin. Navigate to the second "Model" window (*figure 5*). The X and Y coordinates of each origin appear in the table and drop-down menus, where TRX refers to the translation in X and TRY refers to the translation in Y.

Select the "Normal" shape function and open the "Options" dialog (*figure 6*) to specify the "Centre" or mean and "Spread" or standard deviation of each coordinate dimension. The "Centre" parameter represents the average position of the centroid in X and Y and the "Spread" parameter represents the uncertainty (in translation) of that centroid. In this example, assume that the "Centre" value is equal to the 'measured' value of the centroid and the "Spread" is 10 metres. To implement these assumptions via the "Options" dialog (and assuming that the imported object was named "Object1"):

1) Select `Object1_POSITION_TRX_Centre` in the left table and `Object1_POSITION_TRX` in the right table and click "Apply". This assigns the X coordinate of the measured centroid to the "Centre" parameter of the translation in X;
2) repeat (1) for the centre parameter of the Y coordinate (`Object1_POSITION_TRY_Centre`), assigning the measured Y centroid (`Object1_POSITION_TRY`) to that parameter;
3) select `Object1_POSITION_TRX_Spread` and edit the functional relation to read `Object1_POSITION_TRX_Spread = 10`. Click "Apply" to assign a value of 10 to the Spread parameter of the translation in X; and
4) repeat (1) for the Spread parameter of the translation in Y (`Object1_POSITION_TRY_Spread`).

Close the "Options" dialog and validate the model parameters by clicking "**Validate**". Click "**Next**" to enter the first correlation dialog (*figure 7*).

The uncertainty model for a single point in space or time comprises a (marginal) uncertainty model for each coordinate dimension (e.g. X and Y for 2D spatial data), together with any relationships between them. In DUE, these relationships can only be defined for uncertainties that are assumed joint normally distributed, and are then completely specified by a matrix of correlation coefficients. Here, positive correlation will lead to a similar movement in each coordinate dimension. In some cases, an assumption of statistical independence (zero correlation) is appropriate, for which any marginal probability distribution can be applied in DUE (e.g. the Uniform distribution). In many cases, however, an assumption of statistical independence is

unrealistic, because the instruments used to collect positional information or digitise geographic coordinates lead to consistent positional errors.

In this example, the translation in X and Y of the 27 'buildings' requires $3n^2$-3n = 2,106 correlation coefficients (compared to 135,468 as a 'deformable' object). In order to simplify the problem of specifying these correlations, an assumption of 'second-order stationarity' is often made. Here, the correlations depend only on the Euclidean distance between points (and possibly direction), for which a stationary function is assigned. Currently, this is a necessary assumption in DUE. In future, it will be possible to load a custom matrix of correlation coefficients.

In this example, the correlations include the relationships between points in each coordinate dimension (autocorrelations in X and autocorrelations in Y) and the relationships between points across the coordinate dimensions (cross-correlations between X and Y). In the first correlation dialog (*figure 7*), Specify a "correlogram" model for each of the correlation options (autocorrelation and crosscorrelation), as shown in *figure 8*. In selecting cross-correlations, a cross-correlation model must be assigned for all pairs of coordinate attributes (except the rotation coordinates of rigid objects which are excluded for simplicity). Thus, three correlation functions are required here. In practice, it is not straightforward to define a valid correlation matrix when multiple attributes are cross-correlated. One approach to building a valid matrix, often used in spatial statistics, is to specify a set of linearly-related correlation functions; the so-called 'linear model of co-regionalisation' (LMC). This is a strong assumption and is not necessary in DUE, but will produce a valid matrix. The LMC requires that the auto- and cross-correlation functions all comprise the same basic shapes (e.g. Exponential). The LMC is assumed here.

Navigate to the window for defining correlograms (*figure 9*). The two autocorrelation functions, one for a translation in X and one for a translation in Y, appear in a drop-down menu. Select the "`POSITION_TRX`" function and add a single "Exponential" shape to the list. Enter 500m for the "Range" (i.e. an average correlation length of 500m), and click "**Set**" to save the model. Select the "`POSITION_TRY`" function from the drop-down menu and apply the same model and "Range" value.

Finally, select the "`POSITION_TRX_POSITION_TRY`" cross-correlation function from the "Cross-corr." drop-down menu (*figure 7*) and add an "Exponential" shape to the list. For this function, set the "Sill" or maximum cross-correlation to 0.8, and apply a "Range" of 500m (click "**Set**" to store the function). In order to generate a valid correlation matrix, the cross-correlations must be less than the square root of the product of the two autocorrelations at each lag distance (the so-called Cauchy-Schwartz condition), hence the maximum correlation of 0.8. In this case, the LMC has been adhered to and the overall correlation matrix will be valid. Any co-located

points are removed from the correlation matrix before simulation in order to ensure a valid matrix. Click "**Validate**" to check and save the model. The model is now complete and ready for use in a Monte-Carlo study.

Navigate to the "Output" window of DUE (*figure 8*). Simulating a vector object is basically the same as simulating other types of object in DUE (see *Section 5.5*). Specify the number of realisations to produce and the directory to which they should be written. Click "**Run**" to generate the realisations (currently in ESRI shape format only for polygons). Open the realisations in an external data viewer (such as Landserf: www.landserf.org). Notice the similar directions in which (nearby) buildings move in each realisation, reflecting the auto- and cross-correlations between the translations in X and Y.

When defining pdfs for 'deformable objects' that contain overlapping boundaries (duplicate points), such as field boundaries, the duplicate points may be grouped together, in order to maintain the boundaries when simulating from the pdf. The option to 'group coordinates' (true by default) appears in the first "Model" window and again in the "Output" window (when objects have been assigned 'deformable').

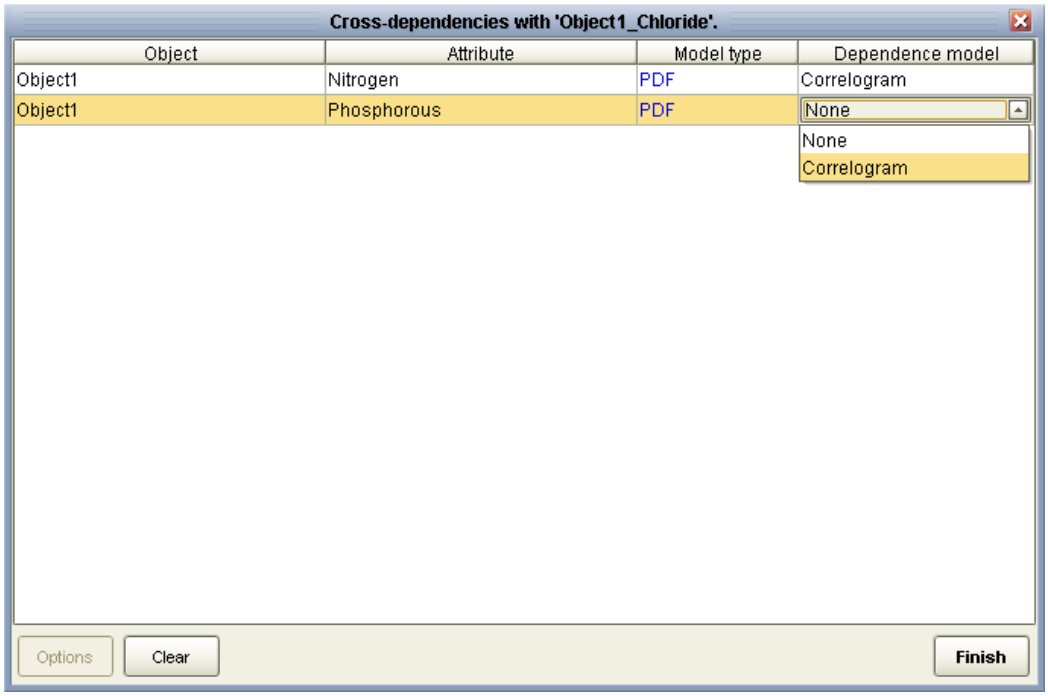
5.8 Co-simulation of multiple, cross-correlated, time-series

Load the "Water_quality.tsd" file from the "due/resources/exampledata" folder in the root directory of your installation (e.g. C:/Program Files/DUE_3.1/due/resources/exampledata). The file contains three water quality time-series, namely Chloride, Nitrogen and Phosphorous, from one chemical monitoring station. Import the attributes into a single object with "POINT_VALUE" for the temporal statistic and "MONTH" for the "Temporal Units" in each case. The attribute units are "MILLIGRAM/LITRE for each of the Chloride, Nitrogen and Phosphorous, attributes. Although the values of different variables are frequently related, correlations between the errors (uncertainties) of multiple variables are less common, as they are typically measured with different equipment. However, a common monitoring station was used to sample C, N and P in this example, which led to consistent uncertainties between attributes.

Co-simulation of multiple cross-correlated attributes in DUE requires the identification of a pdf for each marginal variable, together with the *pairwise* relationships between variables. Thus, for relationships between three or more variables, a full multivariate pdf is constructed iteratively in DUE. Currently, the specification of dependencies between attributes requires a joint normal pdf for each of the dependent variables. Define a normal pdf for each of the Chloride, Nitrogen and Phosphorous attributes, assigning the measured values to the mean and 1.0 for the standard deviation in each case (see Exercise 5.3 first). In addition, specify an autocorrelation model for each attribute, using an exponential shape function with a range of 0.5 months.

After defining the uncertainties of each marginal variable, select the “Chloride” attribute and navigate to the first correlation window (*figure 7*) where the pairwise relationships between attributes are defined. Activate “**Correlated with uncertainties of other attributes**” and then select “Correlogram” for each pair of attributes, as shown in *figure 12*. Click “**OK**” to return to the main window and “**Next**” to enter the correlogram window (*figure 9*).

Figure 12: Defining pairwise relationships between uncertain attributes



As shown in Exercise 5.7, it is not straightforward to define a valid covariance matrix when multiple attributes are cross-correlated. One approach to building a valid matrix, often used in spatial statistics, is to specify a set of linearly-related correlation functions (see above). In this example, one valid matrix is obtained by specifying an exponential shape for all of the autocorrelation functions (above) as well as the cross-correlation functions, together with a range of 0.5 months for each function. The sill of the cross-correlation functions should be 0.5 or less (smaller than the square root of the product of the variances).

Set the autocorrelations for the selected (Chloride) attribute, together with the pairwise correlations between Chloride and Nitrogen and Chloride and Phosphorous. On clicking “**Validate**” the covariance matrices for each of these pairwise relations is constructed and validated. Although all three attributes now appear in the Output window, simulation is restricted to the separate (marginal) attributes or the pairs of attributes for which cross-correlations have been defined. Selecting all three

attributes for simulation will result in a warning message because the pairwise relationship between Nitrogen and Phosphorous has not yet been defined (but Nitrogen and Phosphorous have been implicitly linked through their relationship with Chloride). Define the pairwise relationship between Nitrogen and Phosphorous by selecting the Nitrogen attribute in the Input window (again using an exponential correlation function with a range of 0.5 and a sill of 0.5). All three attributes are now available for co-simulation in the Output window.

APPENDIX A1 CONCEPTUAL BASIS FOR DUE

A1.1 Introduction

Since uncertainty models are influenced by the characteristics of an uncertain variable, it is useful to develop a taxonomy of uncertain environmental variables. The taxonomy is based on objects that may comprise one or more attributes and is used to structure an uncertainty analysis in DUE.

A1.2 Objects and attributes

In this framework, objects are formal descriptions of 'real' entities, and are typically abstractions and simplifications of those entities. Real entities include things with observed boundaries, such as buildings, trees, or storm events, and things with 'fiat' boundaries, such as political borders and calendar years, or some combination of the two (e.g. the Berlin Wall). These boundaries will contain positional information, such as absolute coordinates in space and time or relative distances between locations. If the coordinates or distances are uncertain, the boundaries contain positional uncertainty.

The properties of an object are represented as attributes. In DUE, positional information is represented as one attribute of an object. However, positional uncertainty is distinguished from 'attribute uncertainty' here, as additional simplifications are required for the former. Attribute values may be defined at one or many locations for which the object is defined or described as integral properties of the object. For example a 'river object' may contain the attributes 'length' and 'volume' as integral properties of the object (defined once), together with the attributes 'nutrient concentrations', 'navigation pressures' and 'fish stocks' as distributed properties of the object.

A1.3 Taxonomy of uncertain objects

In order to describe the positional uncertainty of an environmental object, it is useful to classify objects by their primitive parts and by the types of movement they support under uncertainty. A first-order classification would include:

P1. Objects that are single points (point objects);
P2. Objects that comprise multiple points whose relative position in space-time (internal geometry) cannot change under uncertainty (rigid objects);
P3. Objects that comprise multiple points whose relative position in space-time can vary under uncertainty (deformable objects).

In contrast to rigid and deformable objects, the positional uncertainty of a point object always leads to a unitary shift in the object's position. Rigid and deformable objects may comprise groups of isolated points, such as the 'trees' in a 'forest' or the 'animals' in a 'game reserve', groups of interconnected points, such as a 'railway track' or a time series of 'water levels', and closed lines or polygons (in 2D or 3D), such as 'soil mapping units', 'buildings' or 'lakes'. However, the positional uncertainty of a rigid or deformable object is always characterised by the uncertainties of its individual points. The distinction between rigid objects and deformable objects may be physically based if the geometry of an object cannot be altered in principle, or practically motivated if an assumption of rigidity simplifies the pdf.

The positional uncertainty of a rigid object leads to a unitary shift in the object's position (translation) and/or an angular shift (rotation) of the object for any given outcome of the pdf, because the primitive nodes are perfectly correlated. By implication, positional uncertainty cannot alter the topology of a rigid object. In contrast, the topology of a deformable object may be altered by positional uncertainty, because the uncertainties in its primitive points are partially or completely independent of each other.

A1.4 Taxonomy of uncertain attributes

In order to develop probability models for attribute uncertainty, it is useful to distinguish between: 1) the measurement scale of an attribute, and 2) the space-time variability of an attribute (which is partly constrained by the object, unless the object varies in space and time).

Four classes of measurement scale are used in DUE, namely:

1. Attributes measured on a continuous numerical scale (e.g. population density, the diameter of a tree at breast height, annual precipitation);
2 Attributes measured on a discrete numerical scale (e.g. the number of inhabitants in a city or the number of plant species in a forest);
3. Attributes measured on a categorical scale (e.g. soil type or income tax bracket);

In addition, four classes of space-time variability are distinguished, namely:

A. Attributes that are constant in space and time. These include attributes that are known constants, such as the gravitational constant or the universal gas constant, and are effectively certain for environmental research. They also include attributes whose space-time variability is *assumed* constant, such as the threshold at which a chemical concentration leads to fish kills.
B. Attributes that vary in time, but not in space. These include attributes that are

constant in space (e.g. national interest rates in a national economic study) and attributes whose spatial variability is negligible for some practical purpose. In terms of the latter, attributes with a high degree of temporal versus spatial variability might be assumed constant in space for all practical purposes.

C. Attributes that vary in space, but not in time (apply B to time).
D. Attributes that vary in time and space. These include attributes whose temporal variability and spatial variability are both important for some practical application (e.g. precipitation in a global climate study).

The combination of attribute scale (1-3) and space-time variability (A-D) leads to 12 classes of uncertain attributes (*table A1*).

Table A1: Attribute categories for guiding the application of uncertainty models

Space-time variability	**Measurement scale**		
	Continuous numeric	Discrete numeric	Categorical
Constant in space and time	A1	A2	A3
Varies in time, not in space	B1	B2	B3
Varies in space, not in time	C1	C2	C3
Varies in time and space	D1	D2	D3

APPENDIX A2 MODELS AND ALGORITHMS USED IN DUE

A2.1 Introduction

When all possible outcomes of an uncertain event are known and their associated probabilities are quantifiable, uncertainties may be described with a pdf. In order to represent uncertainty with a pdf it is necessary to choose the shape function (assuming the pdf is parametric) and to estimate its parameters at each point in space and time. For objects and attributes that vary in space or time, or for multiple related attributes, the pdf comprises the marginal pdfs (mpdf) at each space time point, together with any correlations between them (see Brown and Heuvelink, 2005 also).

A2.2 Attribute uncertainty

An uncertain continuous numerical constant (or an uncertain variable defined at one point in space and time), is completely specified by its marginal (cumulative) pdf:

$$F_A(a) = P(A \le a) \qquad a \in \Re \qquad (1)$$

The mpdf must be a continuous, non-decreasing, function whose limit values are $F_A(-\infty)=0$ and $F_A(+\infty) = 1$. The corresponding general mpdf for a discrete numerical or categorical attribute is:

$$F_A(a_i) = P(A = a_i) \qquad i = 1, \ldots, n \qquad (2)$$

where the a_i are integers or categories, respectively. Each of the $F_A(a_i)$ should be non-negative and the sum of all $F_A(a_i)$ should be equal to 1. For numerical attributes, most distribution functions F_A have a mean or expected value, $E[A] = \mu_A$, corresponding to the 'bias' of A, and a standard deviation, $\sigma_A = \sqrt{E[(A-\mu_A)^2]}$, corresponding to the 'average uncertainty' of A, both of which are displayed in DUE.

In order to reduce the complexity of an mpdf, the distribution function, F_A, may be described with a simple, parametric, shape. For example, the continuous mpdf in Eqn. 1 may follow a Normal distribution with mean μ and standard deviation σ:

$$F_A(a) = \int_{-\infty}^{a} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx \qquad a \in \Re \qquad (3)$$

Alternatively, a discrete numerical attribute may follow a Poisson distribution with

mean or rate λ:

$$F_A(a) = \frac{e^{-\lambda}\lambda^a}{a!} \qquad a = 1,\ldots,\infty \qquad (4)$$

where $E[A] = \sigma_A = \lambda$. In practice, categorical attributes rarely follow a parametric distribution. In that case, the mpdf, F_A, must be defined for each of the possible outcomes $a_1 \ldots a_n$, as indicated in Eqn. 2. A wide range of parametric distributions is available in DUE, including the Normal, Exponential, Weibull, Beta and Gamma distributions for continuous numerical data, the Poisson, Binomial, Geometric and Bernoulli distributions for discrete numerical data and the discrete Uniform distribution for categorical data (*table A1*).

Table A1: parametric probability models and sampling algorithms used in DUE

Distribution	Sampling method	Reference
Beta	Stratified/patchwork rejection	Sakasegawa (1983); Zechner & Stadlober (1993)
Cauchy	Inversion	Knuth (1998)
ChiSquare	Ratio of uniforms with shift	Monahan (1987)
Cont. Uniform	Mersenne Twister	Matsumoto & Nishimura (1998)
Exponential	Inversion	Knuth (1998)
Gamma	Acceptance/rejection/complement	Ahrens & Dieter (1974,1982)
Gumbel min.	Inversion	Knuth (1998)
Gumbel max.	Inversion	Knuth (1998)
Lognormal	See Normal	-
Normal	Polar method	Knuth (1998)
Triangular	Inversion	Knuth (1998)
Weibull	Inversion	Knuth (1998)
Bernoulli	Compare input with prob. success	-
Binomial	Acceptance/rejection and inversion	Kachitvichyanukul & Schmeiser (1988)
Disc. Uniform	Mersenne Twister	Matsumoto & Nishimura (1998)
Geometric	Inversion	Knuth (1998)
Poisson	Patchwork rejection and inversion	Stadlober & Zechner (1999)

An uncertain continuous numerical variable that varies in one or both of space and time is completely specified by its (cumulative) joint pdf:

$$F_A(a_1, x_1, ..., a_n, x_n) = P(A(x_1) \le a_1, ..., A(x_n) \le a_n) \qquad a, x \in \Re \qquad (5)$$

where the x_n are coordinates and n may assume any integer value. In this context, the "joint pdf" is used to describe a single variable that varies in space or time, and

the “multivariate joint pdf” is used to describe multiple variables that vary jointly in space or time.

The equivalent joint pdf for a discrete numerical or categorical variable is:

$$F_A(a_1, x_1, ..., a_i, x_n) = P(A(x_1) = a_1, ..., A(x_n) = a_i) \qquad a, x \in \Re \qquad (6)$$

where the a_i are integers or categories, respectively, and n may assume any integer value.

The marginal pdfs are obtained from Eqn. 5 by integration. If the mpdfs are statistically independent, the joint pdf is equivalent to the product of the mpdfs. In that case, defining a joint pdf is equivalent to defining an mpdf for each coordinate, x_i, in DUE. If the mpdfs are statistically dependent, the joint pdf includes both the mpdfs and the relationships between them. While numerous parametric models are available for the mpdfs in Eqn. 1 and Eqn. 2, few models are available for the statistically-dependent joint pdf. In the absence of a simple model, the joint probabilities of each combination of a_n and x_n in Eqn. 5 must be defined explicitly. This is prohibitive for variables that occupy more than a few coordinates. Thus, for continuous numerical variables, a common assumption is that Eqn. 5 follows a joint-normal distribution:

$$f_A(x_1, ..., x_n) = \frac{1}{(2\pi)^{n/2} |\sum|^{1/2}} e^{\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right)} \qquad x \in \Re^n \qquad (7)$$

where f_A is the mathematical derivative of F_A with respect to all a_i (i.e., the probability density), n is the number of marginals, μ is a vector of means and Σ is the variance-covariance matrix, which must be symmetric and positive definite. If the latter is satisfied, the determinant of Σ, namely |Σ|, is positive. In assuming Eqn. 7, the pdf is greatly simplified, because it requires only a vector of means and a covariance matrix for complete specification. The joint-normal distribution is currently the only model supported in DUE for statistically dependent mpdfs, with an assumption of statistical independence required in all other cases.

In practice, deriving a realistic and statistically valid (positive definite) covariance matrix is a non-trivial task. A common assumption is that σ is constant for all x_i and that the covariance depends only on the Euclidean distance, |h|, between pairs of x_i, such that $Cov(A(x_i), A(x_j)) = Cov(|h|)$. This is equivalent to deriving Σ from a semivariogram (γ) whereby $Cov(A(x_i), A(x_j)) = \sigma^2 - \gamma(|h|)$. A similar model is available in DUE, except the covariance is derived from ρ, such that

$Cov(A(x_i), A(x_j)) = \sigma^2 \cdot \rho(|h|)$. This allows σ to vary for each x_i while ρ remains a simple function of |h|. DUE supports a wide range of functions of ρ, all of which are proven positive definite, including the exponential, spherical and 'nugget' functions. More complex functions are derived by summing these basic models. For example, the sum of an exponential function and a nugget function leads to an exponential model with a discontinuity at |h| = 0 (a 'nugget effect'). For two- and three-dimensional attributes, ρ can also vary with direction, for which an anisotropy model is used. The model implemented in DUE is equivalent to that in Isaaks and Strivastava (1989) and is not discussed further.

As indicated above, multivariate pdfs are currently only supported for continuous numerical variables. A group of uncertain continuous numerical constants are completely specified by their (cumulative) multivariate pdf:

$$F(a_1,...,a_n) = P(A_1 \le a_1,....,A_n \le a_n) \qquad a_i \in \Re \qquad (8)$$

If the mpdfs in Eqn. 8 are statistically independent, the multivariate pdf is equivalent to the product of the mpdfs. In that case, the multivariate pdf is modelled as a group of mpdfs in DUE, to which separate parametric shapes can be assigned. For the multivariate normal pdf, the cross-correlations between mpdfs are entered manually.

A group of uncertain continuous numerical variables are completely specified by their (cumulative) multivariate joint pdf:

$$F(\mathbf{a}_1, x_1,..., \mathbf{a}_n, x_n) = P(\mathbf{A}_1(x_1) \le \mathbf{a}_1,...., \mathbf{A}_n(x_n) \le \mathbf{a}_n) \qquad a, x \in \Re^n \qquad (9)$$

where each $\mathbf{A}_i$ is p * 1 dimensional vector of random variables at location x_i and n may assume any integer value. As before, an assumption of joint-normality is currently required in DUE if the mpdfs are statistically dependent. In that case, the covariance matrix Σ comprises both the relationships within attributes (autocovariances) and the relationships between attributes (cross-covariances), both of which may vary with x. Four options are available in DUE for specifying the cross-covariances in Σ, namely: 1) statistical independence, such that $Cov(\mathbf{A}_i(x),\mathbf{A}_j(x+h)) = 0$; 2) intrinsic stationarity, such that $Cov(\mathbf{A}_i(x),\mathbf{A}_j(x+h)) = Cov$; 3) second-order stationarity, such that $Cov(\mathbf{A}_i(x),\mathbf{A}_j(x+h)) = \sigma_i \cdot \sigma_j \cdot \rho_{ij}(|h|)$; and 4) an arbitrary positive definite covariance matrix. Although it is not straightforward to derive a valid covariance matrix for the univariate case in Eqn. 5, it is even more complicated for the multivariate case in Eqn. 9. If the vectors of attributes $\mathbf{A}_1,\ldots,\mathbf{A}_n$ are assumed second-order stationary (as in 3 above), a common approach is to invoke the "linear model of co-regionalization", which ensures a positive definite covariance matrix (Goovaerts, 1997). In that case, the cross-covariances are a linear, positive definite,

function of the auto-covariances and are always lower than the square root of the product of the autocovariances at each x (the Cauchy-Schwartz condition). The "linear model of co-regionalization" is not imposed in DUE, but is explained and demonstrated in the user's manual.

A2.3 Positional uncertainty

For simplicity, the coordinate dimensions (xyzt) of an object in DUE, and hence its positional uncertainties, are represented as continuous numerical attributes of that object. Thus, the positional uncertainty of a 'timestamp' is characterised by its marginal pdf in Eqn. 1. Similarly, the positional uncertainty of one location (in space, and possibly time), is characterised by its multivariate pdf in Eqn. 8. Finally, the positional uncertainty of multiple locations in space, and possibly time, are characterised by their multivariate joint pdf in Eqn. 9. The same conditions apply on simplifying the pdf and, given an assumption of normality, on specifying any correlations within and between coordinates. However, in addition to these simplifications, objects that comprise multiple locations in space or time may be classified as 'rigid' or 'deformable' under uncertainty (see above). In this context, a deformable object comprises multiple locations that can move independently, or with partial correlations, under uncertainty. Thus, a deformable object has the same (complex) pdf as a group of continuous numerical variables (i.e. Eqn. 9). In contrast, the pdf of a rigid object comprises the translation (x) and possibly rotation (**θ**) of a single point about that object:

$$F_{X\boldsymbol{\Theta}}(x,\boldsymbol{\theta}) = P(X \le x, \boldsymbol{\Theta} \le \boldsymbol{\theta}) \qquad x,\boldsymbol{\theta} \in \Re \qquad (10)$$

where x is a translation in space and/or time and **θ** is a p * 1 dimensional vector of rotation angles. If x is a four-dimensional space-time coordinate, **θ** contains the three-spatial rotations θ_{XY}, θ_{XZ} and θ_{YZ}, the order of which must be defined (it affects the rotated position), and the three space-time rotations θ_{XT}, θ_{YT} and θ_{ZT}, which are not considered in DUE. In keeping with Eqn. 9, the positional uncertainty of multiple rigid objects is characterised by its multivariate jpdf. Simulation of topologically corrupt objects is prevented in DUE, but may be overridden to simulate complex topologies as (groups of) primitive lines. Sampling of rigid or deformable objects is otherwise identical to the simulation of continuous numerical attributes (see below).

A2.4 Simulation from probability models

For marginal pdfs whose inverse cumulative distribution function (cdf) is available in a simple (analytical) form, a random number is drawn from the mpdf by, first, simulating from a standard Uniform Distribution u ~ U(0,1), and then solving the inverse cdf for u (i.e. the 'inversion method'). Simulation from an mpdf relies on a

pseudorandom number generator that produces uncorrelated random numbers from U(0,1). The "Mersenne Twister" algorithm is used in DUE (Matsumoto and Nishimura, 1998). For distributions whose inverse cdf is not available in an analytical form, distribution specific methods are used to simulate from the mpdf (see *table A1*).

For one or more variables (or multiple constants) whose marginal pdfs are statistically independent, a realisation is drawn from the joint pdf by sampling from the separate mpdfs and pooling the results (*table A1*). As indicated above, the joint-normal distribution is currently the only model supported in DUE for statistically dependent mpdfs. In principle, sampling from the joint normal distribution is straightforward. First, the covariance matrix Σ is factorised to obtain $\sqrt{\sum}$. In DUE, the factorised matrix is obtained from the Cholesky decomposition of Σ. If Σ is a symmetric, positive definite matrix, the Cholesky decomposition is a lower triangular matrix, L, that satisfies:

$$\sum = LL^T \tag{11}$$

where T represents the transpose. Secondly, a vector of samples is obtained from the standard normal distribution N(0,I), with Identity Matrix I, using the Polar method (*table 1*). Sampling from Eqn. 7 then involves rescaling by $\sqrt{\sum}$ (or L), and adding the vector of means μ:

$$x = \mu + L \cdot z \tag{12}$$

where z is a random sample from N(0,I) and x is a random sample from the required distribution, N(μ, Σ). For an attribute with n elements, the covariance matrix will contain n^2 elements. In many cases, Σ is too large to store in memory, or to factorise directly, even in a sparse framework. Hence, the Sequential Simulation Algorithm is used instead of Eqn. 12 for large Σ (Goovaerts, 1997). This relies on the Gstat executable (Pebesma, 2004), which is called through a "command file" for maximum flexibility and portability. In this context, the "platform independence" of DUE is not sacrificed because Gstat is available for all major operating systems.

By linking DUE to Gstat, unconditional and conditional simulations are supported for large Σ. Unconditional simulation is equivalent to sampling from a pdf that was formulated through expert judgement alone. Conditional simulation improves the pdf by combining a model of Σ with direct observations of the uncertain variable(s). In keeping with the assumption of normality, the sample data may be transformed to a Normal distribution. Among others, a 'rank-order transform' is provided in DUE. Here, the observations are transformed to their Normal scores before performing the conditional simulation and back-transformed afterwards (see Goovaerts, 1997).

APPENDIX A3 REFERENCES

Ahrens, J.H. and Dieter, U. (1982) Generating gamma variates by a modified rejection technique. *Communications of the ACM*, 25, 47-54

Ahrens, J.H. and Dieter, U. (1974) Computer methods for sampling from gamma, beta Poisson and binomial distributions. *Computing*, 12, 223-246

Brown, J.D. and Heuvelink, G.B.M (2005) Representing and simulating uncertain environmental variables in GIS. Submitted to *International Journal of Geographical Information Science*

Goovaerts, P. (1997) *Geostatistics for Natural Resources Evaluation*. Oxford University Press: New York

Isaacs, E. H. and Srivastava, R. M. (1989) *An introduction to applied geostatistics*. Oxford University Press, New York

Knuth, D. E. (1998) *The Art of Computer Programming, Vol. 2: Seminumerical Algorithms, 3rd ed.* Addison-Wesley, Reading, MA

Matsumoto, M. and Nishimura, T. (1998) A 623-dimensionally equidistributed uniform pseudorandom number generator. *ACM Transactions on Modeling and Computer Simulation*, 8(1), 3-30

Monahan, J.F. (1987) An algorithm for generating chi random variables. *ACM Transactions of Mathematical Software*, 13, 168-172

Pebesma, E.J. (2004) Multivariable geostatistics in S: the gstat package. *Computers and sciences*, 30, 683-691

Sakasegawa, H. (1983) Stratified rejection and squeeze method for generating beta random numbers. *Annuls of the Institute of Statistical Mathematics*, 35(B), 291-302

Zechner, H. and Stadlober, E. (1993) Generating beta variates via patchwork rejection. *Computing*, 50, 1-18